us-09-873-737a-1.rge

Tue Jul 22 10:58:37 2003

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2003, 18:02:32 ; Search time 7680.12 Seconds
(without alignments)
11546.204 Million cell updates/sec

Title: US-09-873-737A-1
Perfect score: 3047
Sequence: 1 ctgagtccaaagcgtcgttt.....aaaaaaaaaaaaaaaaaaa 3047

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3044	99.9	3047	3	AF104354	AF104354 Drosophil
c 2	1735.4	57.0	78507	2	AC019704	AC019704 Drosophil
c 3	1734.4	56.9	169210	3	AC092226	AC092226 Drosophil
c 4	1734.4	56.9	176306	3	AC007186	AC007186 Drosophil
c 5	1734.4	56.9	276576	3	AE003630	AE003630 Drosophil
6	1717.8	56.4	9359	3	AF104355	AF104355 Drosophil
7	479.4	15.7	2761	3	AF145680	AF145680 Drosophil
8	321.8	10.6	3057	3	AY014899	AY014899 Strongylo
9	314.4	10.3	3910	10	AB032604	AB032604 Mus muscu
10	314.4	10.3	4064	10	AF438405	AF438405 Mus muscu
11	277.6	9.1	2605	3	AY014900	AY014900 Strongylo
12	258.2	8.5	3421	9	AF387507	AF387507 Homo sapi
13	256.6	8.4	2328	9	AF264004	AF264004 Homo sapi
14	253.4	8.3	3362	9	AF104260	AF104260 Homo sapi
15	253.4	8.3	3399	9	BC028581	BC028581 Homo sapi
16	241.8	7.9	3273	5	AF336369	AF336369 Danio rer
17	222.2	7.3	2579	9	AK093133	AK093133 Homo sapi
c 18	221.6	7.3	37980	3	AC005116	AC005116 Drosophil
19	220.2	7.2	901	3	AF334408	AF334408 Drosophil
20	220.2	7.2	3006	9	BC031060	BC031060 Homo sapi
21	218.6	7.2	3138	9	AK093475	AK093475 Homo sapi
22	218	7.2	7781	3	DMDNARPL9	X94613 D.melanogas
23	196.8	6.5	4913	10	AB032605	AB032605 Mus muscu
24	195.2	6.4	2456	9	AK096837	AK096837 Homo sapi
25	168	5.5	1723	3	AY014901	AY014901 Strongylo
26	164	5.4	3649	6	AX247684	AX247684 Sequence
27	164	5.4	3659	10	AF285586	AF285586 Mus muscu
28	156.8	5.1	3610	9	BC025995	BC025995 Homo sapi
29	152.4	5.0	3620	9	AK056418	AK056418 Homo sapi
30	125.8	4.1	2272	9	AK027497	AK027497 Homo sapi
31	123.8	4.1	2845	3	AF247763	AF247763 Drosophil
c 32	114.2	3.7	656	3	AY060859	AY060859 Drosophil
c 33	101.8	3.3	1733	9	AK001213	AK001213 Homo sapi
34	94.6	3.1	3677	2	AC017885	AC017885 Drosophil
c 35	94.6	3.1	3855	3	AE003107	AE003107 Drosophil
36	94.2	3.1	747	3	AF334411S2	AF334412 Drosophil
37	90.4	3.0	153477	2	AC006278	AC006278 Plasmodiu
38	88.4	2.9	113863	2	AC007548	AC007548 Drosophil
c 39	84.4	2.8	27772	2	AC115679	AC115679 Dictyoste
40	84	2.8	1263	9	AK000397	AK000397 Homo sapi
41	83	2.7	2376	2	AC019420	AC019420 Drosophil
c 42	83	2.7	3866	3	AE002585	AE002585 Drosophil
43	82	2.7	68080	2	AC116955	AC116955 Dictyoste
44	81.6	2.7	3392	3	AF300334	AF300334 Dictyoste
c 45	81.4	2.7	10566	3	AC115588	AC115588 Dictyoste

ALIGNMENTS

RESULT 1
AF104354 3047 bp mRNA linear INV 21-JAN-1999
LOCUS AF104354
DEFINITION Drosophila melanogaster PIWI (piwi) mRNA, complete cds.
ACCESSION AF104354
VERSION AF104354.1 GI:4165139
KEYWORDS .
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 3047)
AUTHORS Cox,D.N., Chao,A., Baker,J., Chang,L., Qiao,D. and Lin,H.
TITLE A novel class of evolutionarily conserved genes defined by piwi are

```
  JOURNAL   essential for stem cell self-renewal
  MEDLINE   Genes Dev. 12 (23), 3715-3727 (1998)
   PUBMED   99069219
REFERENCE   9851978
  AUTHORS   2  (bases 1 to 3047)
  TITLE     Cox,D.N. and Lin,H.
  JOURNAL   Direct Submission
            Submitted (04-NOV-1998) Cell Biology, Duke University, Box 3709
            Duke University Medical Center, Durham, NC 27710, USA
FEATURES             Location/Qualifiers
     source          1. .3047
                     /organism="Drosophila melanogaster"
                     /strain="Oregon R"
                     /db_xref="taxon:7227"
                     /chromosome="II"
     gene            /map="32C"
                     1. .3047
     5'UTR           /gene="piwi"
                     1. .83
     CDS             /gene="piwi"
                     84. .2615
                     /gene="piwi"
                     /codon_start=1
                     /product="PIWI"
                     /protein_id="AAD08704.1"
                     /db_xref="GI:4165140"
                     /translation="MADDQGRGRRRPLNEDDSSTSRGSGDGPRVKVFRGSSSGDPRAD
                     PRIEASRERRALEEAPRREGGPPERKPWGDQYDYLNTRPVELVSKKGTDGVPVMLQTN
                     FFRLKTKPEWRIVHYHVEFVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFE
                     QEITVLSGKSKLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLF
                     DPRAKIEIREFKMELWPGYETSIRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPA
                     RHQDEVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNI
                     RIRDHNQPLLISKNRDKALKTNASELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTR
                     MNPKQRTDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKV
                     PAGENADWQRHFRDQRMLTTPSDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRI
                     QVVTLKTTKNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKR
                     AYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIVF
                     YRDGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLSPPQLAYIVVTRSMNTRFFLNGQN
                     RSPQEFIIYDDRTGTYVRAMDDCVRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPT
                     PPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGLSPEKMQKLTYKMC
                     HLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKFYYL"
misc_feature         2301. .2429
                     /gene="piwi"
     3'UTR           /note="encodes PIWI box protein signature"
                     2616. .3047
                     /gene="piwi"
BASE COUNT       936 a    666 c    672 g    773 t
ORIGIN

  Query Match          99.9%;  Score 3044;  DB 3;  Length 3047;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 3044;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 CTGAGTCCAAAGCGTCGTTTTCAAAGTACTCTTTCAGTTTCCATTGTGAAGTTTTAAGTG 60
Db          1 CTGAGTCCAAAGCGTCGTTTTCAAAGTACTCTTTCAGTTTCCATTGTGAAGTTTTAAGTG 60

Qy         61 ATCGCGAGTGCCAAAAAGTAACAATGGCTGATGATCAGGGACGTGGACGCAGGCGTCCAN 120
Db         61 ATCGCGAGTGCCAAAAAGTAACAATGGCTGATGATCAGGGACGTGGACGCAGGCGTCCAC 120

Qy        121 TTAACGAAGATGATTCCTCTACTTCCCGAGGTAGTGGTGATGGGCCGCGGGTGAAAGTAT 180
Db        121 TTAACGAAGATGATTCCTCTACTTCCCGAGGTAGTGGTGATGGGCCGCGGGTGAAAGTAT 180

Qy        181 TCAGAGGATCTTCATCAGGTGACCCGAGAGCGGATCCTCGTATAGAGGCTTCAAGAGAGA 240
Db        181 TCAGAGGATCTTCATCAGGTGACCCGAGAGCGGATCCTCGTATAGAGGCTTCAAGAGAGA 240

Qy        241 GAAGAGCTCTCGAGGAAGCTCCCAGGCGTGAAGGTGGCCCGCCAGAGCGAAAGCCGTGGG 300
Db        241 GAAGAGCTCTCGAGGAAGCTCCCAGGCGTGAAGGTGGCCCGCCAGAGCGAAAGCCGTGGG 300

Qy        301 GTGACCAATATGATTACCTGAATACCCGTCCGGTTGAGCTGGTATCCAAGAAGGGAACCG 360
Db        301 GTGACCAATATGATTACCTGAATACCCGTCCGGTTGAGCTGGTATCCAAGAAGGGAACCG 360

Qy        361 ATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGANTAAAAACCAAGCCGGAATGGC 420
Db        361 ATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGATTAAAAACCAAGCCGGAATGGC 420

Qy        421 GGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAGAATCCTCGTGTCCGTATGG 480
Db        421 GGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAGAATCCTCGTGTCCGTATGG 480

Qy        481 GAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTATCTATTCGACGGACTGCAAC 540
Db        481 GAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTATCTATTCGACGGACTGCAAC 540

Qy        541 TGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTCAGCGGAAAGTCGAAGCTGG 600
Db        541 TGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTCAGCGGAAAGTCGAAGCTGG 600

Qy        601 ACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATATCGTGTGCTGAGCCCCGCT 660
Db        601 ACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATATCGTGTGCTGAGCCCCGCT 660

Qy        661 TTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCGATGAAGGGCCTAAATTTGGAATTAG 720
Db        661 TTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCGATGAAGGGCCTAAATTTGGAATTAG 720

Qy        721 TTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATAAGGGAGTTCAAAATGGAGC 780
Db        721 TTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATAAGGGAGTTCAAAATGGAGC 780

Qy        781 TATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAAGATATTTTATTGGGCACCG 840
Db        781 TATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAAGATATTTTATTGGGCACCG 840

Qy        841 AAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGACATAATGCGACGTTGCTCAC 900
Db        841 AAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGACATAATGCGACGTTGCTCAC 900

Qy        901 ACAATCCGGCTCGTCATCAGGACGAAGTACGGGTAAATGTTTTGGACTTGATTGTCCTTA 960
Db        901 ACAATCCGGCTCGTCATCAGGACGAAGTACGGGTAAATGTTTTGGACTTGATTGTCCTTA 960

Qy        961 CGGATTACAATAACAGAACTTATCGTATCAATGATGTCGACTTTGGACAAACTCCGAAAT 1020
Db        961 CGGATTACAATAACAGAACTTATCGTATCAATGATGTCGACTTTGGACAAACTCCGAAAT 1020

Qy       1021 CAACATTCAGTTGCAAGGGTAGAGATATCAGTTTCGTGGAATACTATCTCACTAAATATA 1080
Db       1021 CAACATTCAGTTGCAAGGGTAGAGATATCAGTTTCGTGGAATACTATCTCACTAAATATA 1080

Qy       1081 ATATACGCATTCGCGACCACAATCAGCCGCTGCTGATTTCCAAAAATAGGGACAAGGCTC 1140
Db       1081 ATATACGCATTCGCGACCACAATCAGCCGCTGCTGATTTCCAAAAATAGGGACAAGGCTC 1140

Qy       1141 TAAAAACTAACGCTAGCGAATTAGTGGTACTAATTCCTGAGCTCTGCCGAGTGACTGGGC 1200
Db       1141 TAAAAACTAACGCTAGCGAATTAGTGGTACTAATTCCTGAGCTCTGCCGAGTGACTGGGC 1200

Qy       1201 TCAATGCCGAGATGCGCTCAAACTTTCAGCTTATGCGTGCCATGAGCAGTTATACGCGAA 1260
Db       1201 TCAATGCCGAGATGCGCTCAAACTTTCAGCTTATGCGTGCCATGAGCAGTTATACGCGAA 1260

Qy       1261 TGAACCCCAAACAACGCACTGATCGATTGCGCGCTTTTAACCACCGTTTACAAAACACTC 1320
Db       1261 TGAACCCCAAACAACGCACTGATCGATTGCGCGCTTTTAACCACCGTTTACAAAACACTC 1320

Qy       1321 CAGAAAGTGTGAAGGTCTTGAGAGACTGGAACATGGAACTGGACAAGAACGTCACAGAAG 1380
Db       1321 CAGAAAGTGTGAAGGTCTTGAGAGACTGGAACATGGAACTGGACAAGAACGTCACAGAAG 1380

Qy       1381 TACAAGGCCGGATAATTGGACAGCAGAACATCGTGTTTCATAATGGAAAGGTTCCTGCTG 1440
```

```
Db      1381 TACAAGGCCGGATAATTGGACAGCAGAACATCGTGTTTCATAATGGAAAGGTTCCTGCTG 1440
Qy      1441 GAGAAAACGCTGATTGGCAAAGGCACTTCAGAGACCAAAGGATGCTTACCACTCCGAGCG 1500
Db      1441 GAGAAAACGCTGATTGGCAAAGGCACTTCAGAGACCAAAGGATGCTTACCACTCCGAGCG 1500
Qy      1501 ATGGCCTCGATCGTTGGGCTGTCATCGCGCCGCAAAGGAATTCCCATGAACTCCGAACTC 1560
Db      1501 ATGGCCTCGATCGTTGGGCTGTCATCGCGCCGCAAAGGAATTCCCATGAACTCCGAACTC 1560
Qy      1561 TACTTGACTCTTTGTATAGAGCAGCTAGTGGAATGGGTCTTAGAATTCGAAGCCCCCAGG 1620
Db      1561 TACTTGACTCTTTGTATAGAGCAGCTAGTGGAATGGGTCTTAGAATTCGAAGCCCCCAGG 1620
Qy      1621 AATTCATAATTTATGATGATCGCACTGGAACTTATGTGAGAGCAATGGATGATTGTGTGC 1680
Db      1621 AATTCATAATTTATGATGATCGCACTGGAACTTATGTGAGAGCAATGGATGATTGTGTGC 1680
Qy      1681 GCTCAGATCCCAAACTTATATTATGCCTCGTACCCAATGATAACGCCGAAAGATACTCAT 1740
Db      1681 GCTCAGATCCCAAACTTATATTATGCCTCGTACCCAATGATAACGCCGAAAGATACTCAT 1740
Qy      1741 CAATCAAAAAGAGAGGATACGTTGACAGGGCGGTGCCAACTCAAGTTGTGACCCTTAAAA 1800
Db      1741 CAATCAAAAAGAGAGGATACGTTGACAGGGCGGTGCCAACTCAAGTTGTGACCCTTAAAA 1800
Qy      1801 CGACCAAGAACCGTAGCCTTATGAGCATTGCCACCAAAATAGCAATCCAACTGAATTGCA 1860
Db      1801 CGACCAAGAACCGTAGCCTTATGAGCATTGCCACCAAAATAGCAATCCAACTGAATTGCA 1860
Qy      1861 AGTTGGGATATACACCCTGGATGATCGAACTACCCTTGTCCGGACTGATGACAATTGGCT 1920
Db      1861 AGTTGGGATATACACCCTGGATGATCGAACTACCCTTGTCCGGACTGATGACAATTGGCT 1920
Qy      1921 TTGACATTGCGAAGAGCACACGAGATCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAA 1980
Db      1921 TTGACATTGCGAAGAGCACACGAGATCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAA 1980
Qy      1981 TGGATCTACAGCAAAACTCCACGTACTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATG 2040
Db      1981 TGGATCTACAGCAAAACTCCACGTACTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATG 2040
Qy      2041 TGCTCGCTAACACCCTTTGGCCGATGATAGCAAAGGCCCTGCGCCAATATCAACATGAGC 2100
Db      2041 TGCTCGCTAACACCCTTTGGCCGATGATAGCAAAGGCCCTGCGCCAATATCAACATGAGC 2100
Qy      2101 ATAGGAAGCTGCCATCTCGAATCGTATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAA 2160
Db      2101 ATAGGAAGCTGCCATCTCGAATCGTATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAA 2160
Qy      2161 AGCAGCTTTTTGAATTTGAAGTCAAGGACATCATTGAGAAGTTGAAAACTGAATACGCCC 2220
Db      2161 AGCAGCTTTTTGAATTTGAAGTCAAGGACATCATTGAGAAGTTGAAAACTGAATACGCCC 2220
Qy      2221 GCGTCCAGCTAAGCCCACCGCAATTAGCTTATATTGTGGTAACCAGATCCATGAACACGC 2280
Db      2221 GCGTCCAGCTAAGCCCACCGCAATTAGCTTATATTGTGGTAACCAGATCCATGAACACGC 2280
Qy      2281 GCTTCTTCCTCAACGGACAAAATCCTCCGCCTGGTACTATAGTTGATGACGTTATAACTC 2340
Db      2281 GCTTCTTCCTCAACGGACAAAATCCTCCGCCTGGTACTATAGTTGATGACGTTATAACTC 2340
Qy      2341 TGCCCGAGAGATACGACTTTTATCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGC 2400
Db      2341 TGCCCGAGAGATACGACTTTTATCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGC 2400
Qy      2401 CGACCAGCTACAATGTTCTTTATAGCAGCATGGGTNTCTCACCGGAGAAAATGCAAAAAC 2460
Db      2401 CGACCAGCTACAATGTTCTTTATAGCAGCATGGGTCTCTCACCGGAGAAAATGCAAAAAC 2460
Qy      2461 TTACGTACAAGATGTGCCACTTGTACTACAATTGGTCGGGCACCACACGAGTGCCAGCAG 2520
Db      2461 TTACGTACAAGATGTGCCACTTGTACTACAATTGGTCGGGCACCACACGAGTGCCAGCAG 2520
Qy      2521 TTTGCCAGTACGCTAAGAAGCTAGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGC 2580
Db      2521 TTTGCCAGTACGCTAAGAAGCTAGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGC 2580
Qy      2581 AAAACGCGCTCGAAAAGAAGTTTTATTATCTATAATTGGATATAATTTAGAATGGAGTAT 2640
Db      2581 AAAACGCGCTCGAAAAGAAGTTTTATTATCTATAATTGGATATAATTTAGAATGGAGTAT 2640
Qy      2641 TAATCCTTACTAAGAGGCCATATATGAAACTAGCCCAGACATTTATACTTTTTCAATACT 2700
Db      2641 TAATCCTTACTAAGAGGCCATATATGAAACTAGCCCAGACATTTATACTTTTTCAATACT 2700
Qy      2701 TCCTTACTTTTGCTAAGCACTTCAGCATTTATGACTAAATATTTTGTATTTGAAATGCAT 2760
Db      2701 TCCTTACTTTTGCTAAGCACTTCAGCATTTATGACTAAATATTTTGTATTTGAAATGCAT 2760
Qy      2761 TACTGCTCTTTTTTCAAACAAAAGCAAAATTGAGGATTAAGATTCTGGTATTTAAGCATA 2820
Db      2761 TACTGCTCTTTTTTCAAACAAAAGCAAAATTGAGGATTAAGATTCTGGTATTTAAGCATA 2820
Qy      2821 AGACCAGAGGAAATTCCCAAACAAACATTTAAAGTTATCTATCAAGACATGTTCATTAAT 2880
Db      2821 AGACCAGAGGAAATTCCCAAACAAACATTTAAAGTTATCTATCAAGACATGTTCATTAAT 2880
Qy      2881 TTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTTTATGTAAAGAAAAATTACA 2940
Db      2881 TTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTTTATGTAAAGAAAAATTACA 2940
Qy      2941 TACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGCTTATATTCGTGAAAAAAAA 3000
Db      2941 TACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGCTTATATTCGTGAAAAAAAA 3000
Qy      3001 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3047
Db      3001 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3047
```

RESULT 2
AC019704/c

```
LOCUS       AC019704               78507 bp    DNA     linear   HTG 03-JAN-2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC019704
VERSION     AC019704.1  GI:6665193
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 78507)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210683 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .78507
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    22865 a  16519 c  16556 g  22567 t
ORIGIN
```

Query Match 57.0%; Score 1735.4; DB 2; Length 78507;
Best Local Similarity 72.7%; Pred. No. 0;
Matches 2845; Conservative 0; Mismatches 48; Indels 1022; Gaps 6;

```
Qy      152 TAGTGGTGATGGGCCGCGGGTGAAAGTATTCAGAGGATCTTCATCAGGTGACCCGAGAGC 211
Db    29995 TAATGCTTTCATGCAGCGGGTGAAAGTATTCAGAGGATCTTCATCAGGTGACCCGAGAGC 29936

Qy      212 GGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGAAGCTCCCAGGCGTGA 271
Db    29935 GGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGAAGCTCCCAGGCGTGA 29876

Qy      272 AGGTGGCCCGCCAGAGCGAAAGCCGTGGGGTGACCAATATGATTACCTGAATACCCGTCC 331
Db    29875 AGGTGGCCCGACAGAGCGAAAGCCGTGGGGTGACCAATATGATTACCTGAATACCCGTCC 29816

Qy      332 GGTTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTT 391
Db    29815 GGCTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTT 29756

Qy      392 TTTTCGANTAAAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCC 451
Db    29755 TTTCCGATTAAAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGAGCC 29696

Qy      452 GACCATCGAGAATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGG 511
Db    29695 GAGCATCGAGAATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGG 29636

Qy      512 ATCAGGCTATCTATTCGACGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAAT 571
Db    29635 ATCAGGCTATCTATTCGATGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAAT 29576

Qy      572 CACGGTGCTCAGCGGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGT 631
Db    29575 CACGGTGCTCAGCGGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGT 29516

Qy      632 TGGATTCATATCGTGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCG 691
Db    29515 TGGATTCATATCGTGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCG 29456

Qy      692 CTCGATGAAGGGCCTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTA- 750
Db    29455 CTCCATGAAGGGCCTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTAA 29396

Qy      751 ----------------------------------------------------------AGATC 755
Db    29395 GGTAATTTCTTTTGGTATCGTTTTAAGTGCTGTTTGGTCAAATATGATTTATTTTAGATC 29336

Qy      756 GAAATAAGGGAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCAC 815
Db    29335 GAAATAAGGGAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCAC 29276

Qy      816 GAAAAAGATATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATC 875
Db    29275 GAAAAAGATATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATC 29216

Qy      876 TACGACATAATGCGACGTTGCTCACACAATCCGGCTCGTCATCAGGACGAAGTACGGGTA 935
Db    29215 TACGACATAATGCGACGTTGCTCACACAATCCGGCTCGTCATCAGGACGAAGTACGGGTA 29156

Qy      936 AATGTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGAT 995
Db    29155 AATGTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGAT 29096

Qy      996 GTCGACTTTGGACAAACTCCGAAATCAACATTCAGTTGCAAGGGTAGAGATATCAGTTTC 1055
Db    29095 GTCGACTTTGGACAAACTCCGAAATCAACATTCAGTTGCAAGGGTAGAGATATCAGTTTC 29036

Qy     1056 GTGGAATACTATCTCACT---------------------------------------- 1073
Db    29035 GTGGAATACTATCTCACTGTAAGTCTAATATTTACCACGGATATACATCGGCGTCTTATT 28976

Qy     1074 ------------------------AAATATAATATACGCATTCGCGACCACAATCAGCCG 1109
Db    28975 TCTCCCAACATTTGTTCTCTTTAGAAATATAATATACGCATTCGCGACCACAATCAGCCG 28916
```

```
Qy     1110 CTGCTGATTTCCAAAAATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTA 1169
Db    28915 CTGCTGATCTCCAAAAATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTA 28856

Qy     1170 CTAATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCA- 1228
Db    28855 CTAATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCAG 28796

Qy     1229 ---------------------------------------------------------- 1228
Db    28795 TAAGACTTTAAACTATATTTAAATTAACAAGCTCTTGTGTCGCAAACTATCTATTTCTTA 28736

Qy     1229 ---------------------------------------------------------- 1228
Db    28735 AGTTATATTCTGGTATCCTTTCTTTACATGACTAAAACTATATTAATTTCGCTAGTCCGT 28676

Qy     1229 ---------------------------------------------------------- 1228
Db    28675 CCGATTCTTGACCTGTTTGGAGTCAACGAGTTCATGCCGATCGGTTCGAAAAAACTTCTT 28616

Qy     1229 ---------------------------------------------------------- 1228
Db    28615 ACTGTTTGCAAACCTGAGATTACCTCCTTATCCCAGTAGACTAATGTTATTACAATAGAA 28556

Qy     1229 ---------------------------------------------------------- 1228
Db    28555 CTAATAACTAATAAGACTAATATGAGATAGAATTATTCCCCGACTATCAGATACCCGTTA 28496

Qy     1229 ---------------------------------------------------------- 1228
Db    28495 CTCAGCTAGTGGAACGGTAATTTTCATCATTTTTCTGTGGTATCGATAGATATTGAGGAA 28436

Qy     1229 ---------------------------------------------------------- 1228
Db    28435 TAAAATGAGAACATTTTTTAAAAATTTAAAACTAAAATTTTAAAAGTCTGGTCGTGACCG 28376

Qy     1229 ---------------------------------------------------------- 1228
Db    28375 TTTTGGCCGTTTCGTATGTGAAATGAGAGGCGCGGCTAAAGTGTTTTAGCTATACAAATA 28316

Qy     1229 ---------------------------------------------------------- 1228
Db    28315 AAAATAGGCAAGACAAACAATAAAACGAAGAAAAATCTGAGCACTTTTCAAAAGTGTGGG 28256

Qy     1229 ---------------------------------------------------------- 1228
Db    28255 CGTGGCAGTTTAGGGCGGTTTTATGGCCCATAAAAAATGTAATATTTAATTATTATAAAC 28196

Qy     1229 --GCTTATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATCGA 1286
Db    28195 AGGCTTATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATCGA 28136

Qy     1287 TTGCGCGCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGAC 1346
Db    28135 TTGCGCGCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGAC 28076

Qy     1347 TGGAACATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCAG 1406
Db    28075 TGGAACATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCAG 28016

Qy     1407 AACATCGTGTTTCATAATGGAAA---------------------------------- 1429
Db    28015 AACATCGTGTTTCATAATGGAAAGTGAGTTGAGCAAGTTAAGAGCACTGCAAGTTTATAG 27956

Qy     1430 ----------------GGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTCAGAG 1473
Db    27955 TTAATAATTTTTTCAGGGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTCAGAG 27896

Qy     1474 ACCAAAGGATGCTTACCACTCCGAGCGATGGCCTCGATCGTTGGGCTGTCATCGCGCCGC 1533
Db    27895 ACCAAAGGATGCTTACCACTCCGAGCGATGGCCTCGATCGTTGGGCTGTCATCGCGCCGC 27836

Qy     1534 AAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAGTGGAA 1593
```

```
Db      27835 AAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAGTGGAA 27776
Qy       1594 TGGGTCTTAGAATTCGAAGCCCCCAGGA---------------------------------- 1621
Db      27775 TGGGTCTTAGAATTCGAAGCCCCCAGGAGTGAGTTTTTATAAAAAAACACGTATACCTCA 27716
Qy       1622 ---------------------------------------------------------- 1621
Db      27715 CATTACGTCTTTACTTAATTCGTAAAGAGAGGTTTCGTGCTTCGTTTTCTGGGTATATAT 27656
Qy       1622 ---------------------------ATTCATAATTTATGATGATCGCACTGGAACTTA 1654
Db      27655 CCACTTTAACTATCGTTTTACAAACAGATTCATAATTTATGATGATCGCACTGGAACTTA 27596
Qy       1655 TGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGCCTCGTACC 1714
Db      27595 TGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGCCTCGTACC 27536
Qy       1715 CAATGATAACGCCGAA---------------------------------------------- 1730
Db      27535 CAATGATAACGCCGAAAGGTATTTATTCCCTAATAAGTGTACAAATCTAAACTTAATGAC 27476
Qy       1731 ---------------------------------------------------------- 1730
Db      27475 TATATTTATCGAAGACATTGTTTAATTTTGGTACACTAAAGTTAATATAAATGAAAGTAC 27416
Qy       1731 ---------------------AGATACTCATCAATCAAAAAGAGAGGATACGTTGACAGG 1769
Db      27415 TGAAGTAAACCTTTTTATTTTAGATACTCATCAATCAAAAAGAGAGGATACGTTGACAGG 27356
Qy       1770 GCGGTGCCAACTCAAGTTGTGACCCTTAAAACGACCAAGAACCGTAGCCTTATGAGCATT 1829
Db      27355 GCGGTGCCAACTCAAGTTGTGACCCTTAAAACGACCAAGAACCGTAGCCTTATGAGCATT 27296
Qy       1830 GCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGATCGAA 1889
Db      27295 GCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGATCGAA 27236
Qy       1890 CTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGATCGG 1949
Db      27235 CTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGATCGG 27176
Qy       1950 AAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTTC 2009
Db      27175 AAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTTC 27116
Qy       2010 AGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATGATA 2069
Db      27115 AGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATGATA 27056
Qy       2070 GCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATTT 2129
Db      27055 GCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATTT 26996
Qy       2130 TATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAAGGAC 2189
Db      26995 TATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAAGGAC 26936
Qy       2190 ATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATTAGCT 2249
Db      26935 ATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATTAGCT 26876
Qy       2250 TATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAATCCTCCG 2309
Db      26875 TATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAATCCTCCG 26816
Qy       2310 CCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTATCTGGTC 2369
Db      26815 CCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTATCTGGTC 26756
Qy       2370 TCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAGCAGC 2429
Db      26755 TCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAGCAGC 26696
Qy       2430 ATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTACTAC 2489
Db      26695 ATGGGTCTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTACTAC 26636
Qy       2490 AATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGCTACC 2549
Db      26635 AATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGCTACC 26576
Qy       2550 CTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTATTAT 2609
Db      26575 CTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTATTAT 26516
Qy       2610 CTATAATTGGATATAATTTAGAATGGAGTATTAATCCTTACTAAGAGGCCATATATGAAA 2669
Db      26515 CTATAATTGGATATAATTTAGAATGGAGTATTAATCCTTACTAAGAGGCCATATATGAAA 26456
Qy       2670 CTAGCCCAGACATTTATACTTTTTCAATACTTCCTTACTTTTGCTAAGCACTTCAGCATT 2729
Db      26455 CTAGCCCAGACATTTATACTTTTTCAATACTTTCTTACTTTTGCTAAGCACTTCAGCATT 26396
Qy       2730 TATGACTAAATATTTTGTATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGCAAAA 2789
Db      26395 TATGACTAAATATTTTGTATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGCAAAA 26336
Qy       2790 TTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAACATT 2849
Db      26335 TTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAACATT 26276
Qy       2850 TAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATTG 2909
Db      26275 TAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATTG 26216
Qy       2910 TATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACAC 2969
Db      26215 TATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACAC 26156
Qy       2970 ATTCATGGCTGCTTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3029
Db      26155 ATTCATGGCTGCTTATATTCGTGAAACATATTTTATTAATTATACAGGTATTTAACTTTC 26096
Qy       3030 AAAAAAAAAAAAAAA 3044
Db      26095 ACAAGGATAATACAA 26081
```

RESULT 3
AC092226/c

```
LOCUS       AC092226              169210 bp    DNA     linear   INV 30-JUN-2001
DEFINITION  Drosophila melanogaster, chromosome 2L, region 32A-32B, BAC clone
            BACR20O15, complete sequence.
ACCESSION   AC092226
VERSION     AC092226.1  GI:14578107
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 169210)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
```

```
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 2L, region 32A-32B
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169210)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUN-2001) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .169210
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
                     /map="32A-32B"
                     /clone="BACR20O15 (D1278)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT    48398 a  36081 c  36539 g  48192 t
ORIGIN

  Query Match             56.9%;  Score 1734.4;  DB 3;  Length 169210;
  Best Local Similarity   72.7%;  Pred. No. 0;
  Matches 2845;  Conservative    0;  Mismatches   48;  Indels 1023;  Gaps    6;

Qy        152 TAGTGGTGATGGGCCGCGGGTGAAAGTATTCAGAGGATCTTCATCAGGTGACCCGAGAGC 211
              || || |    || ||||||||||||||||||||||||||||||||||||||||||||||||
Db      24503 TAATGCTTTCATGCAGCGGGTGAAAGTATTCAGAGGATCTTCATCAGGTGACCCGAGAGC 24444

Qy        212 GGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGAAGCTCCCAGGCGTGA 271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24443 GGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGAAGCTCCCAGGCGTGA 24384

Qy        272 AGGTGGCCCGCCAGAGCGAAAGCCGTGGGGTGACCAATATGATTACCTGAATACCCGTCC 331
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24383 AGGTGGCCCGACAGAGCGAAAGCCGTGGGGTGACCAATATGATTACCTGAATACCCGTCC 24324

Qy        332 GGTTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTT 391
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24323 GGCTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTT 24264

Qy        392 TTTTCGANTAAAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCC 451
              ||| ||| ||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db      24263 TTTCCGATTAAAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGAGCC 24204

Qy        452 GACCATCGAGAATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGG 511
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24203 GAGCATCGAGAATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGG 24144

Qy        512 ATCAGGCTATCTATTCGACGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAAT 571
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      24143 ATCAGGCTATCTATTCGATGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAAT 24084

Qy        572 CACGGTGCTCAGCGGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGT 631
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24083 CACGGTGCTCAGCGGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGT 24024

Qy        632 TGGATTCATATCGTGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCG 691
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24023 TGGATTCATATCGTGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCG 23964

Qy        692 CTCGATGAAGGGCCTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTA- 750
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db      23963 CTCCATGAAGGGCCTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTAA 23904

Qy        751 ----------------------------------------------------------AGATC 755
                                                                           |||||
Db      23903 GGTAATTTCTTTTGGTATCGTTTTAAGTGCTGTTTGGTCAAATATGATTTATTTTAGATC 23844

Qy        756 GAAATAAGGGAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCAC 815
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23843 GAAATAAGGGAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCAC 23784

Qy        816 GAAAAAGATATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATC 875
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23783 GAAAAAGATATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATC 23724

Qy        876 TACGACATAATGCGACGTTGCTCACACAATCCGGCTCGTCATCAGGACGAAGTACGGGTA 935
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23723 TACGACATAATGCGACGTTGCTCACACAATCCGGCTCGTCATCAGGACGAAGTACGGGTA 23664

Qy        936 AATGTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGAT 995
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23663 AATGTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGAT 23604

Qy        996 GTCGACTTTGGACAAACTCCGAAATCAACATTCAGTTGCAAGGGTAGAGATATCAGTTTC 1055
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23603 GTCGACTTTGGACAAACTCCGAAATCAACATTCAGTTGCAAGGGTAGAGATATCAGTTTC 23544

Qy       1056 GTGGAATACTATCTCACT---------------------------------------- 1073
              ||||||||||||||||||
Db      23543 GTGGAATACTATCTCACTGTAAGTCTAATATTTACCACGGATATACATCGGCGTCTTATT 23484

Qy       1074 ------------------------AAATATAATATACGCATTCGCGACCACAATCAGCCG 1109
                                        ||||||||||||||||||||||||||||||||||||
Db      23483 TCTCCCAACATTTGTTCTCTTTAGAAATATAATATACGCATTCGCGACCACAATCAGCCG 23424

Qy       1110 CTGCTGATTTCCAAAAATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTA 1169
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23423 CTGCTGATCTCCAAAAATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTA 23364

Qy       1170 CTAATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCA- 1228
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db      23363 CTAATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCAG 23304

Qy       1229 ---------------------------------------------------------- 1228

Db      23303 TAAGACTTTAAACTATATTTAAATTAACAAGCTCTTGTGTCGCAAACTATCTATTTCTTA 23244

Qy       1229 ---------------------------------------------------------- 1228

Db      23243 AGTTATATTCTGTGTATCCTTTCTTTACATGACTAAAACTATATTAATTTCGCTAGTCCG 23184

Qy       1229 ---------------------------------------------------------- 1228

Db      23183 TCCGATTCTTGACCTGTTTGGAGTCAACGAGTTCATGCCGATCGGTTCGAAAAAACTTCT 23124

Qy       1229 ---------------------------------------------------------- 1228

Db      23123 TACTGTTTGCAAACCTGAGATTACCTCCTTATCCCAGTAGACTAATGTTATTACAATAGA 23064

Qy       1229 ---------------------------------------------------------- 1228
```

```
Db      23063 ACTAATAACTAATAAGACTAATATGAGATAGAATTATTCCCCGACTATCAGATACCCGTT 23004
Qy       1229 ------------------------------------------------------------ 1228
Db      23003 ACTCAGCTAGTGGAACGGTAATTTTCATCATTTTTCTGTGGTATCGATAGATATTGAGGA 22944
Qy       1229 ------------------------------------------------------------ 1228
Db      22943 ATAAAATGAGAACATTTTTTAAAAATTTAAAACTAAAATTTTAAAAGTCTGGTCGTGACC 22884
Qy       1229 ------------------------------------------------------------ 1228
Db      22883 GTTTTGGCCGTTTCGTATGTGAAATGAGAGGCGCGGCTAAAGTGTTTTAGCTATACAAAT 22824
Qy       1229 ------------------------------------------------------------ 1228
Db      22823 AAAAATAGGCAAGACAAACAATAAAACGAAGAAAAATCTGAGCACTTTTCAAAAGTGTGG 22764
Qy       1229 ------------------------------------------------------------ 1228
Db      22763 GCGTGGCAGTTTAGGGCGGTTTTATGGCCCATAAAAAATGTAATATTTAATTATTATAAA 22704
Qy       1229 ---GCTTATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATCG 1285
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22703 CAGGCTTATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATCG 22644
Qy       1286 ATTGCGCGCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGA 1345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22643 ATTGCGCGCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGA 22584
Qy       1346 CTGGAACATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCA 1405
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22583 CTGGAACATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCA 22524
Qy       1406 GAACATCGTGTTTCATAATGGAAA---------------------------------- 1429
              |||||||||||||||||||||||
Db      22523 GAACATCGTGTTTCATAATGGAAAGTGAGTTGAGCAAGTTAAGAGCACTGCAAGTTTATA 22464
Qy       1430 ----------------GGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTCAGA 1472
                            ||||||||||||||||||||||||||||||||||||||||||||
Db      22463 GTTAATAATTTTTTCAGGGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTCAGA 22404
Qy       1473 GACCAAAGGATGCTTACCACTCCGAGCGATGGCCTCGATCGTTGGGCTGTCATCGCGCCG 1532
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22403 GACCAAAGGATGCTTACCACTCCGAGCGATGGCCTCGATCGTTGGGCTGTCATCGCGCCG 22344
Qy       1533 CAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAGTGGA 1592
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22343 CAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAGTGGA 22284
Qy       1593 ATGGGTCTTAGAATTCGAAGCCCCCAGGA---------------------------- 1621
              |||||||||||||||||||||||||||
Db      22283 ATGGGTCTTAGAATTCGAAGCCCCCAGGAGTGAGTTTTTATAAAAAAACACGTATACCTC 22224
Qy       1622 ------------------------------------------------------------ 1621
Db      22223 ACATTACGTCTTTACTTAATTCGTAAAGAGAGGTTTCGTGCTTCGTTTTCTGGGTATATA 22164
Qy       1622 ---------------------------ATTCATAATTTATGATGATCGCACTGGAACTT 1653
                                          |||||||||||||||||||||||||||||||
Db      22163 TCCACTTTAACTATCGTTTTACAAACAGATTCATAATTTATGATGATCGCACTGGAACTT 22104
Qy       1654 ATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGCCTCGTAC 1713
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22103 ATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGCCTCGTAC 22044
Qy       1714 CCAATGATAACGCCGAA---------------------------------------- 1730
              |||||||||||||||||
Db      22043 CCAATGATAACGCCGAAAGGTATTTATTCCCTAATAAGTGTACAAATCTAAACTTAATGA 21984
Qy       1731 ------------------------------------------------------------ 1730
```

```
Db      21983 CTATATTTATCGAAGACATTGTTTAATTTTGGTACACTAAAGTTAATATAAATGAAAGTA 21924
Qy       1731 ----------------------AGATACTCATCAATCAAAAAGAGAGGATACGTTGACAG 1768
                                    ||||||||||||||||||||||||||||||||||||||
Db      21923 CTGAAGTAAACCTTTTTATTTTAGATACTCATCAATCAAAAAGAGAGGATACGTTGACAG 21864
Qy       1769 GGCGGTGCCAACTCAAGTTGTGACCCTTAAAACGACCAAGAACCGTAGCCTTATGAGCAT 1828
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21863 GGCGGTGCCAACTCAAGTTGTGACCCTTAAAACGACCAAGAACCGTAGCCTTATGAGCAT 21804
Qy       1829 TGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGATCGA 1888
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21803 TGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGATCGA 21744
Qy       1889 ACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGATCG 1948
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21743 ACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGATCG 21684
Qy       1949 GAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTT 2008
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21683 GAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTT 21624
Qy       2009 CAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATGAT 2068
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21623 CAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATGAT 21564
Qy       2069 AGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATT 2128
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21563 AGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATT 21504
Qy       2129 TTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAAGGA 2188
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21503 TTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAAGGA 21444
Qy       2189 CATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATTAGC 2248
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21443 CATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATTAGC 21384
Qy       2249 TTATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAATCCTCC 2308
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21383 TTATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAATCCTCC 21324
Qy       2309 GCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTATCTGGT 2368
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21323 GCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTATCTGGT 21264
Qy       2369 CTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAGCAG 2428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21263 CTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAGCAG 21204
Qy       2429 CATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTACTA 2488
              ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21203 CATGGGTCTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTACTA 21144
Qy       2489 CAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGCTAC 2548
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21143 CAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGCTAC 21084
Qy       2549 CCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTATTA 2608
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21083 CCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTATTA 21024
Qy       2609 TCTATAATTGGATATAATTTAGAATGGAGTATTAATCCTTACTAAGAGGCCATATATGAA 2668
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21023 TCTATAATTGGATATAATTTAGAATGGAGTATTAATCCTTACTAAGAGGCCATATATGAA 20964
Qy       2669 ACTAGCCCAGACATTTATACTTTTTCAATACTTCCTTACTTTTGCTAAGCACTTCAGCAT 2728
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      20963 ACTAGCCCAGACATTTATACTTTTTCAATACTTTCTTACTTTTGCTAAGCACTTCAGCAT 20904
Qy       2729 TTATGACTAAATATTTTGTATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGCAAA 2788
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20903 TTATGACTAAATATTTTGTATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGCAAA 20844
```

```
Qy       2789 ATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAACAT 2848
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20843 ATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAACAT 20784

Qy       2849 TTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATT 2908
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20783 TTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATT 20724

Qy       2909 GTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACA 2968
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20723 GTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACA 20664

Qy       2969 CATTCATGGCTGCTTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3028
             |||||||||||||||||||||||||| | |    |  ||  | | |   |   ||
Db      20663 CATTCATGGCTGCTTATATTCGTGAAACATATTTTATTAATTATACAGGTATTTAACTTT 20604

Qy       3029 AAAAAAAAAAAAAAAA 3044
              | ||  | || | ||
Db      20603 CACAAGGATAATACAA 20588
```

RESULT 4
AC007186/c
LOCUS AC007186 176306 bp DNA linear INV 10-MAR-2001
DEFINITION Drosophila melanogaster, chromosome 2L, region 32A-, BAC clone BACR03D06, complete sequence.
ACCESSION AC007186
VERSION AC007186.9 GI:13270528
KEYWORDS HTG.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 176306)
AUTHORS Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A., Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D., Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A., Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J., Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M., Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
TITLE Sequencing of Drosophila chromosome 2L, region 32A-
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 176306)
AUTHORS Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT On Mar 10, 2001 this sequence version replaced gi:7143377.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.fruitfly.org/sequence/) or send email to bdgp@fruitfly.berkeley.edu.
FEATURES Location/Qualifiers
source 1. .176306
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2L"
/map="32A-"
/clone="BACR03D06 (D569)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBACe3.6)"
BASE COUNT 53052 a 36247 c 36094 g 50913 t
ORIGIN

Query Match 56.9%; Score 1734.4; DB 3; Length 176306;
Best Local Similarity 72.7%; Pred. No. 0;
Matches 2845; Conservative 0; Mismatches 48; Indels 1023; Gaps 6;

```
Qy        152 TAGTGGTGATGGGCCGCGGGTGAAAGTATTCAGAGGATCTTCATCAGGTGACCCGAGAGC 211
              || || |     || ||||||||||||||||||||||||||||||||||||||||||||||
Db     155213 TAATGCTTTCATGCAGCGGGTGAAAGTATTCAGAGGATCTTCATCAGGTGACCCGAGAGC 155154

Qy        212 GGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGAAGCTCCCAGGCGTGA 271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     155153 GGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGAAGCTCCCAGGCGTGA 155094

Qy        272 AGGTGGCCCGCCAGAGCGAAAGCCGTGGGGTGACCAATATGATTACCTGAATACCCGTCC 331
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     155093 AGGTGGCCCGACAGAGCGAAAGCCGTGGGGTGACCAATATGATTACCTGAATACCCGTCC 155034

Qy        332 GGTTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTT 391
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     155033 GGCTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTT 154974

Qy        392 TTTTCGANTAAAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCC 451
              ||| ||| ||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     154973 TTTCCGATTAAAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGAGCC 154914

Qy        452 GACCATCGAGAATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGG 511
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     154913 GAGCATCGAGAATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGG 154854

Qy        512 ATCAGGCTATCTATTCGACGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAAT 571
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     154853 ATCAGGCTATCTATTCGATGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAAT 154794

Qy        572 CACGGTGCTCAGCGGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGT 631
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     154793 CACGGTGCTCAGCGGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGT 154734

Qy        632 TGGATTCATATCGTGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCG 691
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     154733 TGGATTCATATCGTGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCG 154674

Qy        692 CTCGATGAAGGGCCTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTA- 750
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     154673 CTCCATGAAGGGCCTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTAA 154614

Qy        751 ----------------------------------------------------------AGATC 755
                                                                         |||||
Db     154613 GGTAATTTCTTTTGGTATCGTTTTAAGTGCTGTTTGGTCAAATATGATTTATTTTAGATC 154554

Qy        756 GAAATAAGGGAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCAC 815
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     154553 GAAATAAGGGAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCAC 154494

Qy        816 GAAAAAGATATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATC 875
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     154493 GAAAAAGATATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATC 154434

Qy        876 TACGACATAATGCGACGTTGCTCACACAATCCGGCTCGTCATCAGGACGAAGTACGGGTA 935
```

```
Db    154433 TACGACATAATGCGACGTTGCTCACACAATCCGGCTCGTCATCAGGACGAAGTACGGGTA 154374
Qy       936 AATGTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGAT 995
Db    154373 AATGTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGAT 154314
Qy       996 GTCGACTTTGGACAAACTCCGAAATCAACATTCAGTTGCAAGGGTAGAGATATCAGTTTC 1055
Db    154313 GTCGACTTTGGACAAACTCCGAAATCAACATTCAGTTGCAAGGGTAGAGATATCAGTTTC 154254
Qy      1056 GTGGAATACTATCTCACT---------------------------------------- 1073
Db    154253 GTGGAATACTATCTCACTGTAAGTCTAATATTTACCACGGATATACATCGGCGTCTTATT 154194
Qy      1074 ------------------------AAATATAATATACGCATTCGCGACCACAATCAGCCG 1109
Db    154193 TCTCCCAACATTTGTTCTCTTTAGAAATATAATATACGCATTCGCGACCACAATCAGCCG 154134
Qy      1110 CTGCTGATTTCCAAAAATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTA 1169
Db    154133 CTGCTGATCTCCAAAAATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTA 154074
Qy      1170 CTAATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCA- 1228
Db    154073 CTAATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCAG 154014
Qy      1229 ---------------------------------------------------------- 1228
Db    154013 TAAGACTTTAAACTATATTTAAATTAACAAGCTCTTGTGTCGCAAACTATCTATTTCTTA 153954
Qy      1229 ---------------------------------------------------------- 1228
Db    153953 AGTTATATTCTGTGTATCCTTTCTTTACATGACTAAAACTATATTAATTTCGCTAGTCCG 153894
Qy      1229 ---------------------------------------------------------- 1228
Db    153893 TCCGATTCTTGACCTGTTTGGAGTCAACGAGTTCATGCCGATCGGTTCGAAAAAACTTCT 153834
Qy      1229 ---------------------------------------------------------- 1228
Db    153833 TACTGTTTGCAAACCTGAGATTACCTCCTTATCCCAGTAGACTAATGTTATTACAATAGA 153774
Qy      1229 ---------------------------------------------------------- 1228
Db    153773 ACTAATAACTAATAAGACTAATATGAGATAGAATTATTCCCCGACTATCAGATACCCGTT 153714
Qy      1229 ---------------------------------------------------------- 1228
Db    153713 ACTCAGCTAGTGGAACGGTAATTTTCATCATTTTTCTGTGGTATCGATAGATATTGAGGA 153654
Qy      1229 ---------------------------------------------------------- 1228
Db    153653 ATAAAATGAGAACATTTTTTAAAAATTTAAAACTAAAATTTTAAAAGTCTGGTCGTGACC 153594
Qy      1229 ---------------------------------------------------------- 1228
Db    153593 GTTTTGGCCGTTTCGTATGTGAAATGAGAGGCGCGGCTAAAGTGTTTTAGCTATACAAAT 153534
Qy      1229 ---------------------------------------------------------- 1228
Db    153533 AAAAATAGGCAAGACAAACAATAAAACGAAGAAAAATCTGAGCACTTTTCAAAAGTGTGG 153474
Qy      1229 ---------------------------------------------------------- 1228
Db    153473 GCGTGGCAGTTTAGGGCGGTTTTATGGCCCATAAAAAATGTAATATTTAATTATTATAAA 153414
Qy      1229 ---GCTTATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATCG 1285
Db    153413 CAGGCTTATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATCG 153354
Qy      1286 ATTGCGCGCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGA 1345
```

```
Db    153353 ATTGCGCGCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGA 153294
Qy       1346 CTGGAACATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCA 1405
Db    153293 CTGGAACATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCA 153234
Qy       1406 GAACATCGTGTTTCATAATGGAAA---------------------------------- 1429
Db    153233 GAACATCGTGTTTCATAATGGAAAGTGAGTTGAGCAAGTTAAGAGCACTGCAAGTTTATA 153174
Qy       1430 ----------------GGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTCAGA 1472
Db    153173 GTTAATAATTTTTTCAGGGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTCAGA 153114
Qy       1473 GACCAAAGGATGCTTACCACTCCGAGCGATGGCCTCGATCGTTGGGCTGTCATCGCGCCG 1532
Db    153113 GACCAAAGGATGCTTACCACTCCGAGCGATGGCCTCGATCGTTGGGCTGTCATCGCGCCG 153054
Qy       1533 CAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAGTGGA 1592
Db    153053 CAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAGTGGA 152994
Qy       1593 ATGGGTCTTAGAATTCGAAGCCCCCAGGA---------------------------- 1621
Db    152993 ATGGGTCTTAGAATTCGAAGCCCCCAGGAGTGAGTTTTTATAAAAAAACACGTATACCTC 152934
Qy       1622 ---------------------------------------------------------- 1621
Db    152933 ACATTACGTCTTTACTTAATTCGTAAAGAGAGGTTTCGTGCTTCGTTTTCTGGGTATATA 152874
Qy       1622 ----------------------------ATTCATAATTTATGATGATCGCACTGGAACTT 1653
Db    152873 TCCACTTTAACTATCGTTTTACAAACAGATTCATAATTTATGATGATCGCACTGGAACTT 152814
Qy       1654 ATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGCCTCGTAC 1713
Db    152813 ATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGCCTCGTAC 152754
Qy       1714 CCAATGATAACGCCGAA---------------------------------------- 1730
Db    152753 CCAATGATAACGCCGAAAGGTATTTATTCCCTAATAAGTGTACAAATCTAAACTTAATGA 152694
Qy       1731 ---------------------------------------------------------- 1730
Db    152693 CTATATTTATCGAAGACATTGTTTAATTTTGGTACACTAAAGTTAATATAAATGAAAGTA 152634
Qy       1731 ----------------------AGATACTCATCAATCAAAAAGAGAGGATACGTTGACAG 1768
Db    152633 CTGAAGTAAACCTTTTTATTTTAGATACTCATCAATCAAAAAGAGAGGATACGTTGACAG 152574
Qy       1769 GGCGGTGCCAACTCAAGTTGTGACCCTTAAAACGACCAAGAACCGTAGCCTTATGAGCAT 1828
Db    152573 GGCGGTGCCAACTCAAGTTGTGACCCTTAAAACGACCAAGAACCGTAGCCTTATGAGCAT 152514
Qy       1829 TGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGATCGA 1888
Db    152513 TGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGATCGA 152454
Qy       1889 ACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGATCG 1948
Db    152453 ACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGATCG 152394
Qy       1949 GAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTT 2008
Db    152393 GAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTT 152334
Qy       2009 CAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATGAT 2068
Db    152333 CAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATGAT 152274
Qy       2069 AGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATT 2128
Db    152273 AGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATT 152214
```

```
Qy       2129 TTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAAGGA 2188
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     152213 TTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAAGGA 152154

Qy       2189 CATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATTAGC 2248
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     152153 CATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATTAGC 152094

Qy       2249 TTATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAATCCTCC 2308
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     152093 TTATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAATCCTCC 152034

Qy       2309 GCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTATCTGGT 2368
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     152033 GCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTATCTGGT 151974

Qy       2369 CTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAGCAG 2428
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151973 CTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAGCAG 151914

Qy       2429 CATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTACTA 2488
             ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151913 CATGGGTCTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTACTA 151854

Qy       2489 CAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGCTAC 2548
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151853 CAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGCTAC 151794

Qy       2549 CCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTATTA 2608
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151793 CCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTATTA 151734

Qy       2609 TCTATAATTGGATATAATTTAGAATGGAGTATTAATCCTTACTAAGAGGCCATATATGAA 2668
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151733 TCTATAATTGGATATAATTTAGAATGGAGTATTAATCCTTACTAAGAGGCCATATATGAA 151674

Qy       2669 ACTAGCCCAGACATTTATACTTTTTCAATACTTCCTTACTTTTGCTAAGCACTTCAGCAT 2728
             |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     151673 ACTAGCCCAGACATTTATACTTTTTCAATACTTTCTTACTTTTGCTAAGCACTTCAGCAT 151614

Qy       2729 TTATGACTAAATATTTTGTATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGCAAA 2788
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151613 TTATGACTAAATATTTTGTATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGCAAA 151554

Qy       2789 ATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAACAT 2848
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151553 ATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAACAT 151494

Qy       2849 TTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATT 2908
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151493 TTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATT 151434

Qy       2909 GTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACA 2968
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151433 GTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACA 151374

Qy       2969 CATTCATGGCTGCTTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3028
             |||||||||||||||||||||||||| | |    |  ||  | | |   |   ||
Db     151373 CATTCATGGCTGCTTATATTCGTGAAACATATTTTATTAATTATACAGGTATTTAACTTT 151314

Qy       3029 AAAAAAAAAAAAAAAA 3044
              | ||  | || | ||
Db     151313 CACAAGGATAATACAA 151298
```

```
RESULT 5
AE003630/c
LOCUS       AE003630              276576 bp    DNA     linear   INV 04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386055 section 23
            of 63, complete sequence.
ACCESSION   AE003630 AE002690
VERSION     AE003630.1  GI:7297758
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 276576)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
   PUBMED   10731132
REFERENCE   2  (bases 1 to 276576)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
FEATURES             Location/Qualifiers
     source          1. .276576
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
     gene            complement(<160. .>1899)
                     /gene="CG6724"
                     /map="32A4-32A5"
                     /db_xref="FLYBASE:FBan0006724"
                     /db_xref="FLYBASE:FBgn0032298"
     mRNA            complement(join(160. .533,601. .847,905. .1571,1632. .1899))
                     /gene="CG6724"
                     /product="CT20871"
                     /db_xref="FLYBASE:FBan0006724"
                     /db_xref="FLYBASE:FBgn0032298"
     CDS             complement(join(244. .533,601. .847,905. .1571,1632. .1690))
                     /gene="CG6724"
                     /note="CG6724 gene product"
                     /codon_start=1
```

```
                     /protein_id="AAF53009.1"
                     /db_xref="GI:7297759"
                     /db_xref="FLYBASE:FBan0006724"
                     /db_xref="FLYBASE:FBgn0032298"
                     /translation="MDVDNGEGQVQVHLKTKQEHYAVPDVPYAIDGTVTTVELNTFVN
                     ALLRQKDGSSDTDFDFLVFDEYLRGRLCDHLREKAISFEDAIEIEYVERFPAPEPQDC
                     LLHDDWVSAVKARGKWILSGCYDNTLNLWTNKGKHILTISGHTAPIKAVDWISLDEET
                     GRFVSTSQDQTAMLWQWNVGSNSVECVSVCKGHERGVDSVSVSPDGLRFATGSWDTML
                     KVWSAELDDAVEGSSKRMKESGVRTPKITLQGHRESVSAVQWMDATTLLTGSWDHTLK
                     VWDLSLEGIKTEISTNKSIFDASYSKLNRLILTASADKNLRLYDPRTNQGSVVRNTYL
                     GHNAWVQTVMWSTTEEFLFVSGAYDNQNKLWDCRSPKAPLYDLLGHGEKVLDIDWSNP
                     KYIVSGGVDNTVRVFKSRKALAEEAETK"
gene                 <2942. .>3238
                     /gene="CG17127"
                     /evidence=not_experimental
                     /map="32A4-32A5"
                     /db_xref="FLYBASE:FBan0017127"
                     /db_xref="FLYBASE:FBgn0032299"
mRNA                 <2942. .>3238
                     /gene="CG17127"
                     /product="CT33650"
                     /evidence=not_experimental
                     /db_xref="FLYBASE:FBan0017127"
                     /db_xref="FLYBASE:FBgn0032299"
CDS                  2942. .3238
                     /gene="CG17127"
                     /note="CG17127 gene product"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAF53010.1"
                     /db_xref="GI:7297760"
                     /db_xref="FLYBASE:FBan0017127"
                     /db_xref="FLYBASE:FBgn0032299"
                     /translation="MKVFALCSMLALLLAGAQSLPQNREGAAYTNEAIRQAQQTLLIP
                     KDAQIQNVQEGIELGAYEQIPGNQRINLFEILGDQVPSEVINNLQSQVDQIGRN"
gene                 complement(<4992. .>7693)
                     /gene="CG18499"
                     /evidence=not_experimental
                     /map="32A4-32A5"
                     /db_xref="FLYBASE:FBan0018499"
                     /db_xref="FLYBASE:FBgn0032300"
mRNA                 complement(join(<4992. .6622,6696. .6797,7156. .7243,
                     7322. .7435,7531. .7693))
                     /gene="CG18499"
                     /product="CT38026"
                     /evidence=not_experimental
                     /db_xref="FLYBASE:FBan0018499"
                     /db_xref="FLYBASE:FBgn0032300"
CDS                  complement(join(6752. .6797,7156. .7243,7322. .7435,
                     7531. .7693))
                     /gene="CG18499"
                     /note="CG18499 gene product"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAF53011.1"
                     /db_xref="GI:7297761"
                     /db_xref="FLYBASE:FBan0018499"
                     /db_xref="FLYBASE:FBgn0032300"
                     /translation="MESSSASSAVVTVAVTPSILSSSSSGGSVGGFHGLAASSTESPS
                     SASPEQQQPQGQCLQDGLWVAESATRCQYQTSCRAIQRTGHCCPDYKCDCEKDGKTYA
                     NGDKLVAPDTPCTVCYCKGERRDCLQPGDLLPAR"
gene                 complement(<20685. .>33042)
                     /gene="CG17130"
                     /map="32A5-32B1"
                     /db_xref="FLYBASE:FBan0017130"
                     /db_xref="FLYBASE:FBgn0032302"
mRNA                 complement(join(20685. .20934,32629. .>33042))
                     /gene="CG17130"
                     /product="CT33644"
                     /db_xref="FLYBASE:FBan0017130"
                     /db_xref="FLYBASE:FBgn0032302"
CDS                  complement(20685. .20927)
                     /gene="CG17130"
                     /note="CG17130 gene product"
                     /codon_start=1
                     /protein_id="AAF53012.1"
                     /db_xref="GI:7297762"
                     /db_xref="FLYBASE:FBan0017130"
                     /db_xref="FLYBASE:FBgn0032302"
                     /translation="MHRTEKHSEAFQLRVSPEFHRKDSGTDLQMSRKKCNNNQQRRGR
                     RRRTDLGNPLGIHGSISLAILMLTICMANTSTAGKC"
gene                 complement(<24291. .>24569)
                     /gene="CG18561"
                     /evidence=not_experimental
                     /map="32A5-32B1"
                     /db_xref="FLYBASE:FBan0018561"
                     /db_xref="FLYBASE:FBgn0032301"
mRNA                 complement(<24291. .>24569)
                     /gene="CG18561"
                     /product="CT42416"
                     /evidence=not_experimental
                     /db_xref="FLYBASE:FBan0018561"
                     /db_xref="FLYBASE:FBgn0032301"
CDS                  complement(24291. .24569)
                     /gene="CG18561"
                     /note="CG18561 gene product"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAF53013.1"
                     /db_xref="GI:7297763"
                     /db_xref="FLYBASE:FBan0018561"
                     /db_xref="FLYBASE:FBgn0032301"
                     /translation="MPKKMNQHSERSTYHVQEWHAIEDAASASASAPAPASSDSVDDD
                     DGAADGQMKMATCGHGSINTKTSSQWAMHRKRLLKQTKVPARSPSGED"
gene                 complement(<36226. .>64053)
                     /gene="Nos"
                     /note="CG6713"
                     /map="32B1-32B3"
                     /db_xref="FLYBASE:FBan0006713"
                     /db_xref="FLYBASE:FBgn0011676"
mRNA                 complement(join(36226. .36583,36787. .36853,37576. .37700,
                     37942. .39084,39248. .39581,44244. .44311,44421. .44525,
                     45258. .45402,47981. .48190,48920. .49101,50349. .50523,
                     54064. .54203,54925. .55066,55718. .55809,57631. .57793,
                     57906. .58042,62557. .62671,63286. .64053))
                     /gene="Nos"
                     /product="CT20853"
                     /note="Nucleotide sequence of the Celera sequence differs
                     from the published sequence for this transcript."
                     /db_xref="FLYBASE:FBan0006713"
                     /db_xref="FLYBASE:FBgn0011676"
CDS                  complement(join(36476. .36583,36787. .36853,37576. .37700,

 Query Match             56.9%;  Score 1734.4;  DB 3;  Length 276576;
 Best Local Similarity   72.7%;  Pred. No. 0;
 Matches 2845;  Conservative    0;  Mismatches   48;  Indels 1023;  Gaps    6;

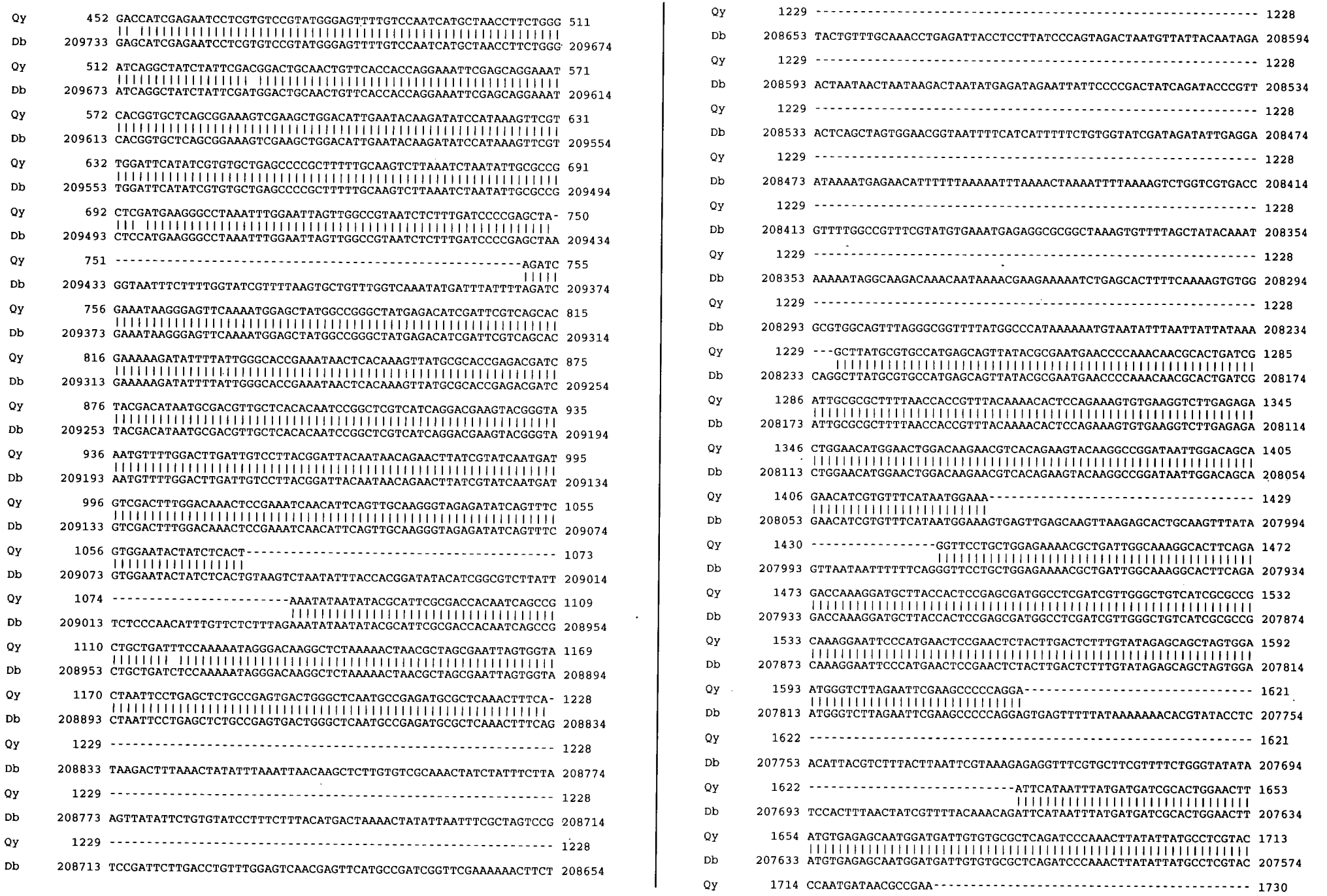
Qy        152 TAGTGGTGATGGGCCGCGGGTGAAAGTATTCAGAGGATCTTCATCAGGTGACCCGAGAGC 211
              || || |     || ||||||||||||||||||||||||||||||||||||||||||||||
Db     210033 TAATGCTTTCATGCAGCGGGTGAAAGTATTCAGAGGATCTTCATCAGGTGACCCGAGAGC 209974

Qy        212 GGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGAAGCTCCCAGGCGTGA 271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     209973 GGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGAAGCTCCCAGGCGTGA 209914

Qy        272 AGGTGGCCCGCCAGAGCGAAAGCCGTGGGGTGACCAATATGATTACCTGAATACCCGTCC 331
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     209913 AGGTGGCCCGACAGAGCGAAAGCCGTGGGGTGACCAATATGATTACCTGAATACCCGTCC 209854

Qy        332 GGTTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTT 391
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     209853 GGCTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTT 209794

Qy        392 TTTTCGANTAAAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCC 451
              ||| ||| ||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     209793 TTTCCGATTAAAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGAGCC 209734
```


Qy 452 GACCATCGAGAATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGG 511
Db 209733 GAGCATCGAGAATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGG 209674
Qy 512 ATCAGGCTATCTATTCGACGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAAT 571
Db 209673 ATCAGGCTATCTATTCGATGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAAT 209614
Qy 572 CACGGTGCTCAGCGGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGT 631
Db 209613 CACGGTGCTCAGCGGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGT 209554
Qy 632 TGGATTCATATCGTGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCG 691
Db 209553 TGGATTCATATCGTGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCG 209494
Qy 692 CTCGATGAAGGGCCTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTA- 750
Db 209493 CTCCATGAAGGGCCTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTAA 209434
Qy 751 ----------AGATC 755
Db 209433 GGTAATTTCTTTTGGTATCGTTTTAAGTGCTGTTTGGTCAAATATGATTTATTTTAGATC 209374
Qy 756 GAAATAAGGGAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCAC 815
Db 209373 GAAATAAGGGAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCAC 209314
Qy 816 GAAAAAGATATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATC 875
Db 209313 GAAAAAGATATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATC 209254
Qy 876 TACGACATAATGCGACGTTGCTCACACAATCCGGCTCGTCATCAGGACGAAGTACGGGTA 935
Db 209253 TACGACATAATGCGACGTTGCTCACACAATCCGGCTCGTCATCAGGACGAAGTACGGGTA 209194
Qy 936 AATGTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGAT 995
Db 209193 AATGTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGAT 209134
Qy 996 GTCGACTTTGGACAAACTCCGAAATCAACATTCAGTTGCAAGGGTAGAGATATCAGTTTC 1055
Db 209133 GTCGACTTTGGACAAACTCCGAAATCAACATTCAGTTGCAAGGGTAGAGATATCAGTTTC 209074
Qy 1056 GTGGAATACTATCTCACT---------- 1073
Db 209073 GTGGAATACTATCTCACTGTAAGTCTAATATTTACCACGGATATACATCGGCGTCTTATT 209014
Qy 1074 ----------AAATATAATATACGCATTCGCGACCACAATCAGCCG 1109
Db 209013 TCTCCCAACATTTGTTCTCTTTAGAAATATAATATACGCATTCGCGACCACAATCAGCCG 208954
Qy 1110 CTGCTGATTTCCAAAAATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTA 1169
Db 208953 CTGCTGATCTCCAAAAATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTA 208894
Qy 1170 CTAATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCA- 1228
Db 208893 CTAATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCAG 208834
Qy 1229 ---------- 1228
Db 208833 TAAGACTTTAAACTATATTTAAATTAACAAGCTCTTGTGTCGCAAACTATCTATTTCTTA 208774
Qy 1229 ---------- 1228
Db 208773 AGTTATATTCTGTGTATCCTTTCTTTACATGACTAAAACTATATTAATTTCGCTAGTCCG 208714
Qy 1229 ---------- 1228
Db 208713 TCCGATTCTTGACCTGTTTGGAGTCAACGAGTTCATGCCGATCGGTTCGAAAAAACTTCT 208654
Qy 1229 ---------- 1228
Db 208653 TACTGTTTGCAAACCTGAGATTACCTCCTTATCCCAGTAGACTAATGTTATTACAATAGA 208594
Qy 1229 ---------- 1228
Db 208593 ACTAATAACTAATAAGACTAATATGAGATAGAATTATTCCCCGACTATCAGATACCCGTT 208534
Qy 1229 ---------- 1228
Db 208533 ACTCAGCTAGTGGAACGGTAATTTTCATCATTTTTCTGTGGTATCGATAGATATTGAGGA 208474
Qy 1229 ---------- 1228
Db 208473 ATAAAATGAGAACATTTTTTAAAAATTTAAAACTAAAATTTTAAAAGTCTGGTCGTGACC 208414
Qy 1229 ---------- 1228
Db 208413 GTTTTGGCCGTTTCGTATGTGAAATGAGAGGCGCGGCTAAAGTGTTTTAGCTATACAAAT 208354
Qy 1229 ---------- 1228
Db 208353 AAAAATAGGCAAGACAAACAATAAAACGAAGAAAAATCTGAGCACTTTTCAAAAGTGTGG 208294
Qy 1229 ---------- 1228
Db 208293 GCGTGGCAGTTTAGGGCGGTTTTATGGCCCATAAAAAATGTAATATTTAATTATTATAAA 208234
Qy 1229 ---GCTTATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATCG 1285
Db 208233 CAGGCTTATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATCG 208174
Qy 1286 ATTGCGCGCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGA 1345
Db 208173 ATTGCGCGCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGA 208114
Qy 1346 CTGGAACATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCA 1405
Db 208113 CTGGAACATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCA 208054
Qy 1406 GAACATCGTGTTTCATAATGGAAA---------- 1429
Db 208053 GAACATCGTGTTTCATAATGGAAAGTGAGTTGAGCAAGTTAAGAGCACTGCAAGTTTATA 207994
Qy 1430 ----------GGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTCAGA 1472
Db 207993 GTTAATAATTTTTTCAGGGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTCAGA 207934
Qy 1473 GACCAAAGGATGCTTACCACTCCGAGCGATGGCCTCGATCGTTGGGCTGTCATCGCGCCG 1532
Db 207933 GACCAAAGGATGCTTACCACTCCGAGCGATGGCCTCGATCGTTGGGCTGTCATCGCGCCG 207874
Qy 1533 CAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAGTGGA 1592
Db 207873 CAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAGTGGA 207814
Qy 1593 ATGGGTCTTAGAATTCGAAGCCCCCAGGA---------- 1621
Db 207813 ATGGGTCTTAGAATTCGAAGCCCCCAGGAGTGAGTTTTTATAAAAAAACACGTATACCTC 207754
Qy 1622 ---------- 1621
Db 207753 ACATTACGTCTTTACTTAATTCGTAAAGAGAGGTTTCGTGCTTCGTTTTCTGGGTATATA 207694
Qy 1622 ----------ATTCATAATTTATGATGATCGCACTGGAACTT 1653
Db 207693 TCCACTTTAACTATCGTTTTACAAACAGATTCATAATTTATGATGATCGCACTGGAACTT 207634
Qy 1654 ATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGCCTCGTAC 1713
Db 207633 ATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGCCTCGTAC 207574
Qy 1714 CCAATGATAACGCCGAA---------- 1730

```
Db     207573 CCAATGATAACGCCGAAAGGTATTTATTCCCTAATAAGTGTACAAATCTAAACTTAATGA 207514
Qy       1731 ------------------------------------------------------------ 1730
Db     207513 CTATATTTATCGAAGACATTGTTTAATTTTGGTACACTAAAGTTAATATAAATGAAAGTA 207454
Qy       1731 ----------------------AGATACTCATCAATCAAAAAGAGAGGATACGTTGACAG 1768
Db     207453 CTGAAGTAAACCTTTTTATTTTAGATACTCATCAATCAAAAAGAGAGGATACGTTGACAG 207394
Qy       1769 GGCGGTGCCAACTCAAGTTGTGACCCTTAAAACGACCAAGAACCGTAGCCTTATGAGCAT 1828
Db     207393 GGCGGTGCCAACTCAAGTTGTGACCCTTAAAACGACCAAGAACCGTAGCCTTATGAGCAT 207334
Qy       1829 TGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGATCGA 1888
Db     207333 TGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGATCGA 207274
Qy       1889 ACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGATCG 1948
Db     207273 ACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGATCG 207214
Qy       1949 GAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTT 2008
Db     207213 GAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTT 207154
Qy       2009 CAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATGAT 2068
Db     207153 CAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATGAT 207094
Qy       2069 AGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATT 2128
Db     207093 AGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATT 207034
Qy       2129 TTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAAGGA 2188
Db     207033 TTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAAGGA 206974
Qy       2189 CATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATTAGC 2248
Db     206973 CATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATTAGC 206914
Qy       2249 TTATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAATCCTCC 2308
Db     206913 TTATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAATCCTCC 206854
Qy       2309 GCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTATCTGGT 2368
Db     206853 GCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTATCTGGT 206794
Qy       2369 CTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAGCAG 2428
Db     206793 CTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAGCAG 206734
Qy       2429 CATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTACTA 2488
Db     206733 CATGGGTCTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTACTA 206674
Qy       2489 CAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGCTAC 2548
Db     206673 CAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGCTAC 206614
Qy       2549 CCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTATTA 2608
Db     206613 CCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTATTA 206554
Qy       2609 TCTATAATTGGATATAATTTAGAATGGAGTATTAATCCTTACTAAGAGGCCATATATGAA 2668
Db     206553 TCTATAATTGGATATAATTTAGAATGGAGTATTAATCCTTACTAAGAGGCCATATATGAA 206494
Qy       2669 ACTAGCCCAGACATTTATACTTTTTCAATACTTCCTTACTTTTGCTAAGCACTTCAGCAT 2728
Db     206493 ACTAGCCCAGACATTTATACTTTTTCAATACTTTCTTACTTTTGCTAAGCACTTCAGCAT 206434
Qy       2729 TTATGACTAAATATTTTGTATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGCAAA 2788
Db     206433 TTATGACTAAATATTTTGTATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGCAAA 206374
Qy       2789 ATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAACAT 2848
Db     206373 ATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAACAT 206314
Qy       2849 TTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATT 2908
Db     206313 TTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATT 206254
Qy       2909 GTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACA 2968
Db     206253 GTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACA 206194
Qy       2969 CATTCATGGCTGCTTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3028
Db     206193 CATTCATGGCTGCTTATATTCGTGAAACATATTTTATTAATTATACAGGTATTTAACTTT 206134
Qy       3029 AAAAAAAAAAAAAAAA 3044
Db     206133 CACAAGGATAATACAA 206118
```

```
RESULT 6
AF104355
LOCUS       AF104355               9359 bp    DNA     linear   INV 21-JAN-1999
DEFINITION  Drosophila melanogaster PIWI (piwi) gene, complete cds.
ACCESSION   AF104355
VERSION     AF104355.1  GI:4165141
KEYWORDS    .
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 9359)
  AUTHORS   Cox,D.N., Chao,A., Baker,J., Chang,L., Qiao,D. and Lin,H.
  TITLE     A novel class of evolutionarily conserved genes defined by piwi are
            essential for stem cell self-renewal
  JOURNAL   Genes Dev. 12 (23), 3715-3727 (1998)
  MEDLINE   99069219
   PUBMED   9851978
REFERENCE   2  (bases 1 to 9359)
  AUTHORS   Cox,D.N. and Lin,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1998) Cell Biology, Duke University, Box 3709
            Duke University Medical Center, Durham, NC 27710, USA
FEATURES             Location/Qualifiers
     source          1. .9359
                     /organism="Drosophila melanogaster"
                     /strain="Canton S"
                     /db_xref="taxon:7227"
                     /chromosome="II"
                     /map="32C"
     gene            3722. .8874
                     /gene="piwi"
     mRNA            join(3722. .3888,5021. .5605,5662. .5982,6049. .6203,
                     6809. .7009,7063. .7254,7374. .7484,7611. .8874)
                     /gene="piwi"
                     /product="piwi"
     CDS             join(3805. .3888,5021. .5605,5662. .5982,6049. .6203,
                     6809. .7009,7063. .7254,7374. .7484,7611. .8496)
                     /gene="piwi"
                     /codon_start=1
                     /product="PIWI"
                     /protein_id="AAD08705.1"
                     /db_xref="GI:4165142"
                     /translation="MADDQGRGRRRPLNEDDSSTSRGSGDGPRVKVFRGSSSGDPRAD
                     PRIEASRERRALEEAPRREGGPTERKPWGDQYDYLNTRPAELVSKKGTDGVPVMLQTN
```

```
                     FFRLKTKPEWRIVHYHVEFEPSIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFE
                     QEITVLSGKSKLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLF
                     DPRAKIEIREFKMELWPGYETSIRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPA
                     RHQDEVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNI
                     RIRDHNQPLLISKNRDKALKTNASELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTR
                     MNPKQRTDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKV
                     PAGENADWQRHFRDQRMLTTPSDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRI
                     RSPQEFIIYDDRTGTYVRAMDDCVRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPT
                     QVVTLKTTKKPYSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRK
                     RAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIV
                     FYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLSPPQLAYIVVTRSMNTRFFLNGQ
                     NPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGLSPEKMQKLTYKM
                     CHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKFYYL"
     misc_feature    8188. .8316
                     /gene="piwi"
                     /note="encodes PIWI box protein signature"
BASE COUNT     2743 a   1839 c   1922 g   2824 t     31 others
ORIGIN

  Query Match           56.4%;  Score 1717.8;  DB 3;  Length 9359;
  Best Local Similarity  72.5%;  Pred. No. 0;
  Matches 2841;  Conservative    4;  Mismatches   48;  Indels 1028;  Gaps    7;

Qy       152 TAGTGGTGATGGGCCGCGGGTGAAAGTATTCAGAGGATCTTCATCAGGTGACCCGAGAGC 211
            || || |     || ||||||||||||||||||||||||||||||||||||||||||||||
Db      5005 TAATGCTTTCATGCAGCGGGTGAAAGTATTCAGAGGATCTTCATCAGGTGACCCGAGAGC 5064

Qy       212 GGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGAAGCTCCCAGGCGTGA 271
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5065 GGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGAAGCTCCCAGGCGTGA 5124

Qy       272 AGGTGGCCCGCCAGAGCGAAAGCCGTGGGGTGACCAATATGATTACCTGAATACCCGTCC 331
            |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5125 AGGTGGCCCGACAGAGCGAAAGCCGTGGGGTGACCAATATGATTACCTGAATACCCGTCC 5184

Qy       332 GGTTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTT 391
            || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5185 GGCTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTT 5244

Qy       392 TTTTCGANTAAAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCC 451
            ||| ||| ||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db      5245 TTTCCGATTAAAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGAGCC 5304

Qy       452 GACCATCGAGAATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGG 511
            || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5305 GAGCATCGAGAATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGG 5364

Qy       512 ATCAGGCTATCTATTCGACGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAAT 571
            ||||||||||||||||||:||:||||||||||||||||||||||||||||||||||||||
Db      5365 ATCAGGCTATCTATTCGAYGGRCTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAAT 5424

Qy       572 CACGGTGCTCAGCGGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGT 631
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5425 CACGGTGCTCAGCGGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGT 5484

Qy       632 TGGATTCATATCGTGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCG 691
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5485 TGGATTCATATCGTGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCG 5544

Qy       692 CTCGATGAAGGGCCTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTA- 750
            ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5545 CTCCATGAAGGGCCTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTAA 5604

Qy       751 ------------------------------------------------------AGATC 755
                                                                     |||||
Db      5605 GGTAATTTCTTTTGGTATCGTTTTAAGTGCTGTTTGGTCAAATATGATTTATTTTAGATC 5664

Qy       756 GAAATAAGGGAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCAC 815
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5665 GAAATAAGGGAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCAC 5724

Qy       816 GAAAAAGATATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATC 875
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      5725 GAAAAAGATATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATC 5784

Qy       876 TACGACATAATGCGACGTTGCTCACACAATCCGGCTCGTCATCAGGACGAAGTACGGGTA 935
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5785 TACGACATAATGCGACGTTGCTCACACAATCCGGCTCGTCATCAGGACGAAGTACGGGTA 5844

Qy       936 AATGTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGAT 995
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5845 AATGTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGAT 5904

Qy       996 GTCGACTTTGGACAAACTCCGAAATCAACATTCAGTTGCAAGGGTAGAGATATCAGTTTC 1055
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5905 GTCGACTTTGGACAAACTCCGAAATCAACATTCAGTTGCAAGGGTAGAGATATCAGTTTC 5964

Qy      1056 GTGGAATACTATCTCACT---------------------------------------- 1073
            ||||||||||||||||||
Db      5965 GTGGAATACTATCTCACTGTAAGTCTAATATTTACCACGGATATACATCGGCGTCTTATT 6024

Qy      1074 -----------------------AAATATAATATACGCATTCGCGACCACAATCAGCCG 1109
                                    ||||||||||||||||||||||||||||||||||||
Db      6025 TCTCCCAACATTTGTTCTCTTTAGAAATATAATATACGCATTCGCGACCACAATCAGCCG 6084

Qy      1110 CTGCTGATTTCCAAAAATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTA 1169
            ||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6085 CTGCTGATYTCCAAAAATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTA 6144

Qy      1170 CTAATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCA- 1228
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6145 CTAATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCAG 6204

Qy      1229 ---------------------------------------------------------- 1228

Db      6205 TAAGACTNTNGACNATATTTAAATTAACAAGCTCTTGTGTCGCAAACTATCTATTTCTTA 6264

Qy      1229 ---------------------------------------------------------- 1228

Db      6265 AGTTATATTCTGTGTATCCTTTCTTATACATGACTAAAACTATATTAATTTCGCTAGTCC 6324

Qy      1229 ---------------------------------------------------------- 1228

Db      6325 CGTCCGATTCTAGACCTGTTTGGAGTCAACGAGTTCAGCCGATCGGTTCGAAAAAACTTC 6384

Qy      1229 ---------------------------------------------------------- 1228

Db      6385 TTANTGTTTGCAAACGTGAGATTACNTCCTTATCCCAGTAGACTAATGTTATTACGATAG 6444

Qy      1229 ---------------------------------------------------------- 1228

Db      6445 NACTNATNACTAATAAGACTAATATGAGATAGAATTATTCCCCGAGTATCAGATACCCGT 6504

Qy      1229 ---------------------------------------------------------- 1228

Db      6505 TACTCAGCTAGTGGAACGGTAATTTTCATCATTTTTCTGTGGTATCGATAGATATTGAGG 6564

Qy      1229 ---------------------------------------------------------- 1228

Db      6565 AATAAAATGAGAACATTTTTTAAAAATTTAAAACTAAAATTTTNAAAGTCTGGTCGTGAC 6624

Qy      1229 ---------------------------------------------------------- 1228

Db      6625 CGTTTTGGCCGTTTCGTATGTGAAATGAGAGGCGCGGCTAAAGTGTTTTAGCTATACAAA 6684

Qy      1229 ---------------------------------------------------------- 1228

Db      6685 TAAAAATAGGCAAGACAAACAATAANACGAAGAAAAATCTGAGCACTTTTCAAAAGTGTG 6744

Qy      1229 ---------------------------------------------------------- 1228

Db      6745 GGCGTGGCAGTTTAGGGCGGTTTTATGGCCCATAAAAAATGTAATATTTAATTATTATAA 6804

Qy      1229 ----GCTTATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATC 1284
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6805 ACAGGCTTATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATC 6864
```

```
Qy       1285 GATTGCGCGCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAG 1344
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6865 GATTGCGCGCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAG 6924

Qy       1345 ACTGGAACATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGC 1404
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6925 ACTGGAACATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGC 6984

Qy       1405 AGAACATCGTGTTTCATAATGGAAA---------------------------------- 1429
             |||||||||||||||||||||||||
Db       6985 AGAACATCGTGTTTCATAATGGAAAGTGAGTTGAGCAAGTTAAGAGCACTGCAAGTTTAT 7044

Qy       1430 ------------------GGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTCAG 1471
                               ||||||||||||||||||||||||||||||||||||||||||||
Db       7045 AGTTAATAATTTTTTCAGGGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTCAG 7104

Qy       1472 AGACCAAAGGATGCTTACCACTCCGAGCGATGGCCTCGATCGTTGGGCTGTCATCGCGCC 1531
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7105 AGACCAAAGGATGCTTACCACTCCGAGCGATGGCCTCGATCGTTGGGCTGTCATCGCGCC 7164

Qy       1532 GCAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAGTGG 1591
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7165 GCAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAGTGG 7224

Qy       1592 AATGGGTCTTAGAATTCGAAGCCCCCAGGA------------------------------ 1621
             ||||||||||||||||||||||||||||
Db       7225 AATGGGTCTTAGAATTCGAAGCCCCCAGGAGTGAGTTTTTATAAAAAAACACGTATACCT 7284

Qy       1622 ------------------------------------------------------------ 1621

Db       7285 CACATTACGTCTTTACTTAATTCGTAAAGAGAGGTTTCGTGCTTCGTTTTCTGGGTATAT 7344

Qy       1622 ----------------------------ATTCATAATTTATGATGATCGCACTGGAACT 1652
                                            ||||||||||||||||||||||||||||||
Db       7345 ATCCACTTTAACTATCGTTTTACAAACAGATTCATAATTTATGATGATCGCACTGGAACT 7404

Qy       1653 TATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGCCTCGTA 1712
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7405 TATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGCCTCGTA 7464

Qy       1713 CCCAATGATAACGCCGAA---------------------------------------- 1730
             ||||||||||||||||||
Db       7465 CCCAATGATAACGCCGAAAGGTATTTATTCCCTAATAAGTGTACAAATCTAAACTTAATG 7524

Qy       1731 ------------------------------------------------------------ 1730

Db       7525 ACTATATTTATCGAAGACATTGTTTAATTTTGGTACACTAAAGTTAATATNAATGAAAGT 7584

Qy       1731 ------------------------AGATACTCATCAATCAAAAAGAGAGGATACGTTGAC 1766
                                       ||||||||||||||||||||||||||||||||||||
Db       7585 NCTGAAGTAAACCTTTNTATTTTTAGATACTCATCAATCAAAAAGAGAGGATACGTTGAC 7644

Qy       1767 AGGGCGGTGCCAACTCAAGTTGTGACCCTTAAAACGACCAAGAACC---GTAGCCTTATG 1823
             |||||||||||||||||||||||||||||||||||||||||||||   |     |||||||||
Db       7645 AGGGCGGTGCCAACTCAAGTTGTGACCCTTAAAACGACCAAGAAACCGTACAGCCTTATG 7704

Qy       1824 AGCATTGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATG 1883
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7705 AGCATTGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATG 7764

Qy       1884 ATCGAACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGA 1943
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7765 ATCGAACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGA 7824

Qy       1944 GATCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACG 2003
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7825 GATCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACG 7884

Qy       2004 TACTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCG 2063
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7885 TACTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCG 7944

Qy       2064 ATGATAGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATC 2123
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7945 ATGATAGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATC 8004

Qy       2124 GTATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTC 2183
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8005 GTATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTC 8064

Qy       2184 AAGGACATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAA 2243
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8065 AAGGACATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAA 8124

Qy       2244 TTAGCTTATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAAT 2303
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8125 TTAGCTTATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAAT 8184

Qy       2304 CCTCCGCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTAT 2363
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8185 CCTCCGCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTAT 8244

Qy       2364 CTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTAT 2423
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8245 CTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTAT 8304

Qy       2424 AGCAGCATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTG 2483
             |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       8305 AGCAGCATGGGTCTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTG 8364

Qy       2484 TACTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTA 2543
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8365 TACTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTA 8424

Qy       2544 GCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTT 2603
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8425 GCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTT 8484

Qy       2604 TATTATCTATAATTGGATATAATTTAGAATGGAGTATTAATCCTTACTAAGAGGCCATAT 2663
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8485 TATTATCTATAATTGGATATAATTTAGAATGGAGTATTAATCCTTACTAAGAGGCCATAT 8544

Qy       2664 ATGAAACTAGCCCAGACATTTATACTTTTTCAATACTTCCTTACTTTTGCTAAGCACTTC 2723
             ||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db       8545 ATGAAACTAGCCCAGACATTTATACTTTTTCAATACTTYCTTACTTTTGCTAAGCACTTC 8604

Qy       2724 AGCATTTATGACTAAATATTTTGTATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAA 2783
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8605 AGCATTTATGACTAAATATTTTGTATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAA 8664

Qy       2784 GCAAAATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCAAACA 2843
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8665 GCAAAATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCAAACA 8724

Qy       2844 AACATTTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTT 2903
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8725 AACATTTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTT 8784

Qy       2904 TTATTGTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAA 2963
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8785 TTATTGTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAA 8844

Qy       2964 CCACACATTCATGGCTGCTTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3023
             |||||||||||||||||||||||||||||||| | |    | ||  | | |   |  |
Db       8845 CCACACATTCATGGCTGCTTATATTCGTGAAACATATTTTATTAATTATACAGGTATTTA 8904

Qy       3024 AAAAAAAAAAAAAAAAAAAAA 3044
             |    | ||  | || | ||
Db       8905 ACTTTCACAAGGATAATACAA 8925
```

RESULT 7
AF145680
LOCUS AF145680 2761 bp mRNA linear INV 19-JUL-1999
DEFINITION Drosophila melanogaster LD23107 sting (sting) mRNA, complete cds.

```
ACCESSION   AF145680
VERSION     AF145680.1  GI:5052649
KEYWORDS    .
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 2761)
  AUTHORS   Tsang,G., Brokstein,P., Frise,E., Harvey,D., Evans-Holm,M.,
            Lewis,S.E., Suh,C. and Rubin,G.M.
  TITLE     Full Length Drosophila melanogaster cDNA sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2761)
  AUTHORS   Rubin,G.M., Wan,K.H., Harvey,D., Lewis,S.E., Brokstein,P.,
            Tsang,G., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Doyle,C.M.,
            Farfan,D.E., Frise,E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Evans-Holm,M., Houston,K.A., Hummasti,S.R., Kim,E.,
            Li,P., Moshrefi,M., Pacleb,J.M., Park,S., Sequeira,A., Sethi,H.,
            Snir,E., Svirskas,R.R., Weinburg,T. and Celniker,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-1999) Berkeley Drosophila Genome Project,
            University of California Berkeley, Berkeley, CA 94720, USA
REFERENCE   3  (bases 1 to 2761)
  AUTHORS   Tsang,G., Brokstein,P., Frise,E., Harvey,D., Evans-Holm,M.,
            Lewis,S.E., Suh,C. and Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-1999) Berkeley Drosophila Genome Project,
            University of California Berkeley, 539 Life Sciences Addition
            #3200, Berkeley, CA 94720, USA
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            University of California Berkeley
            Berkeley, CA 94720
            For further information about this sequence,including its location
            and relationship to other sequences, please visit our Web site
            (http://fruitfly.berkeley.edu) or send email to
            cdna@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .2761
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LD23107"
     gene            1. .2761
                     /gene="sting"
     CDS             29. .2629
                     /gene="sting"
                     /codon_start=1
                     /product="sting"
                     /protein_id="AAD38655.1"
                     /db_xref="GI:5052650"
                     /db_xref="FLYBASE:FBcd0023107"
                     /db_xref="FLYBASE:FBgn0024835"
                     /translation="MNLPPNPVIARGRGRGRKPNNVEANRGFAPSLGQKSDPSHSEGN
                     QASGGNGGGGDAQVGPSIEKSSLSAVQMHKSEGDPRGSVRGRRLITDLVYSRPPGMTS
                     KKGVVGTHITVQANYFKVLKRPNWTIYQYRVDFTPDVEATRLRRSFLYEHKGILGGYI
                     FDGTNMFCINQFKAVQDSPYVLELVTKSRAGENIEIKIKAVGSVQSTDAEQFQVLNLI
                     LRRAMEGLDLKLVSRYYYDPQAKINLENFRMQLWPGYQTSIRQHENDILLCSEICHKV
                     MRTETLYNILSDAIRDSDDYQSTFKRAVMGMVILTDYNNKTYRIDDVDFQSTPLCKFK
                     TNDGEISYVDYYKKRYNIIIRDLKQPLVMSRPTDKNIRGGNDQAIMIIPELARATGMT
                     DAMRADFRTLRAMSEHTRLNPDRRIERLRMFNKRLKSCKQSVETLKSWNIELDSALVE
                     IPARVLPPEKILFGNQKIFVCDARADWTNEFRTCSMFKNVHINRWYVITPSRNLRETQ
                     EFVQMCIRTASSMKMNICNPIYEEIPDDRNGTYSQAIDNAAANDPQIVMVVMRSPNEE
                     KYSCIKKRTCVDRPVPSQVVTLKVIAPRQQKPTGLMSIATKVVIQMNAKLMGAPWQVV
                     IPLHGLMTVGFDVCHSPKNKNKSYGAFVATMDQKESFRYFSTVNEHIKGQELSEQMSV
                     NMACALRSYQEQHRSLPERILFFRDGVGDGQLYQVVNSEVNTLKDRLDEIYKSAGKQE
                     GCRMTFIIVSKRINSRYFTGHRNPVPGTVVDDVITLPERYDFFLVSQAVRIGTVSPTS
                     YNVISDNMGLNADKLQMLSYKMTHMYYNYSGTIRVPAVCHYAHKLAFLVAESINRAPS
                     AGLQNQLYFL"
BASE COUNT      776 a    637 c    663 g    685 t
ORIGIN
```

```
Query Match           15.7%;  Score 479.4;  DB 3;  Length 2761;
Best Local Similarity  52.1%;  Pred. No. 3.3e-93;
Matches 1263;  Conservative    0;  Mismatches 1113;  Indels   47;  Gaps    7;

Qy       312 GATTACCTGAATACCCGTCCGGTTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTCCCG 371
            ||||  || || | |||||    | | ||     |||||| |||   | ||||   |
Db       299 GATTTGGTGTATTCACGTCCTCCAGGGATGACCAGCAAGAAAGGAGTGGTTGGCACTCAT 358

Qy       372 GTCATGCTGCAGACGAACTTTTTTCGANTAAAAACCAAGCCGGAATGGCGGATCGTTCAT 431
              |  ||||| | |||| ||||    ||  ||      ||  | |||   |||   ||
Db       359 ATTACCGTGCAGGCAAACTATTTTAAGGTATTAAAGCGTCCAAACTGGACCATCTACCAG 418

Qy       432 TATCACGTGGAGTTTGTGCCGACCATCGAGAATCCTCGTGTCCGTATGGGAGTTTTGTCC 491
            || | ||| || |||  |||     | |||  | ||  | ||    |   | ||||
Db       419 TACCGCGTCGATTTTACGCCTGATGTGGAGGCTACACGACTGCGACGGTCTTTCTTGTAT 478

Qy       492 AATCATGCTAACCTTCTGGGATCAGGCTATCTATTCGACGGACTGCAACTGTTCACCACC 551
              | |||       | |||||    |||||  | || ||||||    |  ||||   || |
Db       479 GAACATAAAGGGATCCTGGG---CGGCTACATCTTTGACGGAACCAATATGTTTTGCATC 535

Qy       552 AGGAAATTCGAGCAGG---------------AAATCACGGTGCTCAGCGGAAAGTCGAAG 596
              |  | ||| |   |                |  |  ||  ||       |||
Db       536 AATCAGTTCAAAGCTGTTCAAGATAGCCCCTATGTTTTGGAACTGGTTACGAAGAGTCGT 595

Qy       597 CTGGACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATATCGTGTGCTGAGCCC 656
              | |   |     | ||   || |||   |||||||   |    | | | ||  ||
Db       596 GCTGGCGAAAACATTGAAATTAAGATCAAGGCTGTTGGATCTGTGCAATCTACGGATGCC 655

Qy       657 CGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCGATGAAGGGCCTAAATTTGGAA 716
              || || ||| | ||||| ||| ||||| |  | ||| |||||| || | ||| |
Db       656 GAGCAGTTTCAGGTCCTTAATCTCATACTGCGCAGGGCCATGGAGGGCTTAGACTTGAAG 715

Qy       717 TTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATAAGGGAGTTCAAAATG 776
              | ||    ||  |   ||  ||||| | |||||||||  |  |   | | |||   |||
Db       716 CTGGTCTCGCGCTACTACTACGATCCTCAAGCTAAGATTAATTTGGAGAATTTCCGCATG 775

Qy       777 GAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAAGATATTTTATTGGGC 836
              |  ||||||| |||||| |||| |||||||| |||||||||| ||||| ||| |  |
Db       776 CAATTATGGCCTGGCTATCAGACTTCGATTCGCCAGCACGAAAATGATATATTACTTTGT 835

Qy       837 ACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGACATAATGCGACGTTGC 896
              |||| |||   ||||| |||||||| || |||||  | ||| | ||  ||     |
Db       836 TCCGAGATATGCCACAAGGTTATGCGAACTGAGACCTTGTACAATATTTTGTCCGATGCT 895

Qy       897 TCACACAATCCGGCTCGTCATCAGGACGAAGTACGGGTAAATGTTTTGGACTTGATTGTC 956
              | | |    |   | ||||     | |   |       ||  |||   || | |
Db       896 ATTCGCGACAGTGACGATTATCAAAGTACATTTAAGCGTGCAGTAATGGGTATGGTAATA 955

Qy       957 CTTACGGATTACAATAACAGAACTTATCGTATCAATGATGTCGACTTTGGACAAACTCCG 1016
             || |||||||| ||||||| ||| ||||| ||  | ||||||||||||  |   || ||
Db       956 CTAACGGATTATAATAACAAAACCTATCGCATTGACGATGTCGACTTTCAATCGACGCCA 1015

Qy      1017 AAATCAACATTCAGTTGCAAGGGTAGAGATATCAGTTTCGTGGAATACTATCTCACTAAA 1076
              |  | ||| |   ||| |   | || ||    | |||||| ||||||   |   |
Db      1016 TTGTGCAAATTTAAAACCAATGACGGTGAGATTTCGTACGTGGATTACTATAAGAAGCGA 1075

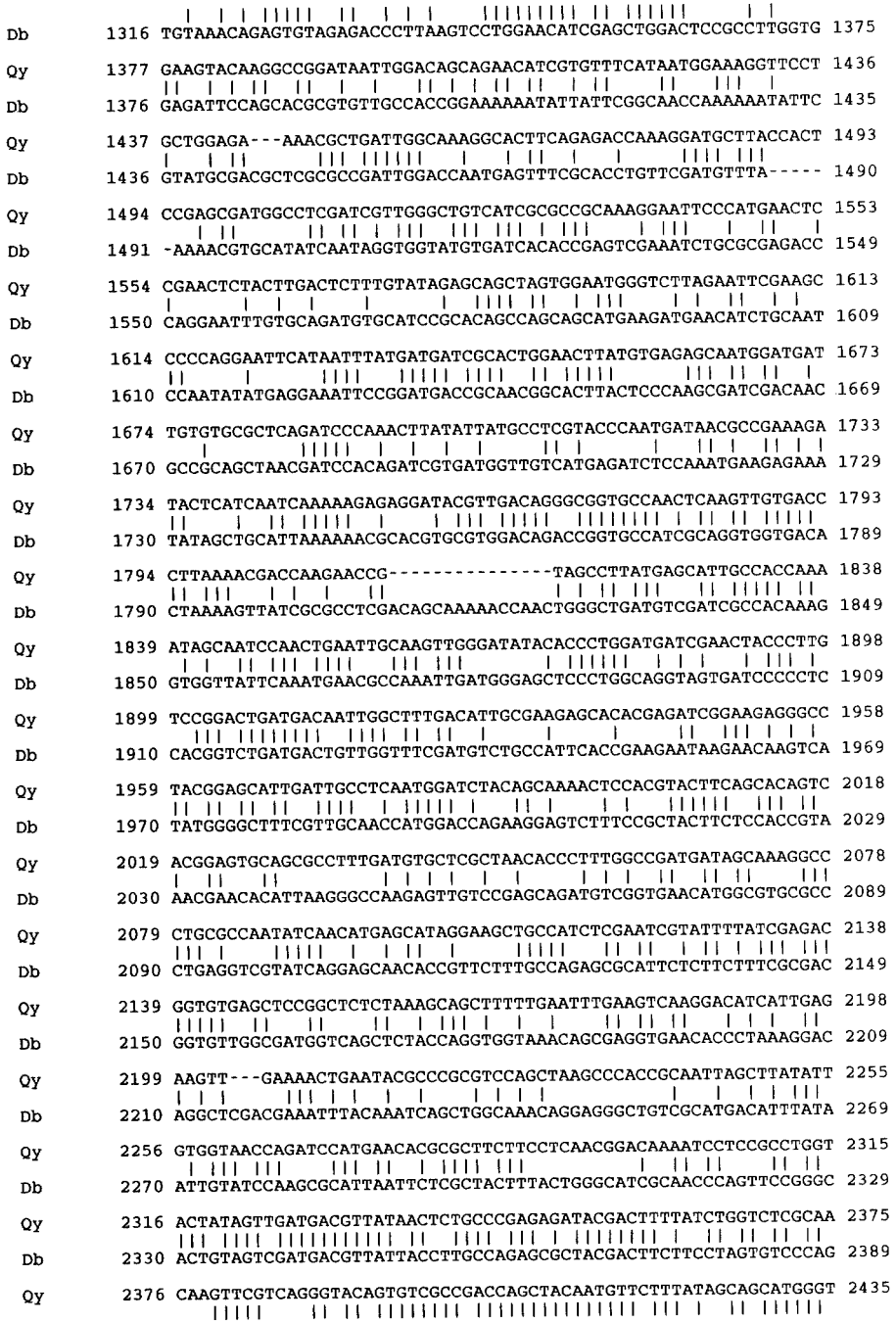
Qy      1077 TATAATATACGCATTCGCGACCACAATCAGCCGCTGCTGATTTCCAAAAATAGGGACAAG 1136
             || || |||  |||||||||||  || || || ||| | || ||       |  |||||
Db      1076 TACAACATAATCATTCGCGACTTGAAGCAACCTCTGGTCATGTCTCGTCCGACAGACAAA 1135

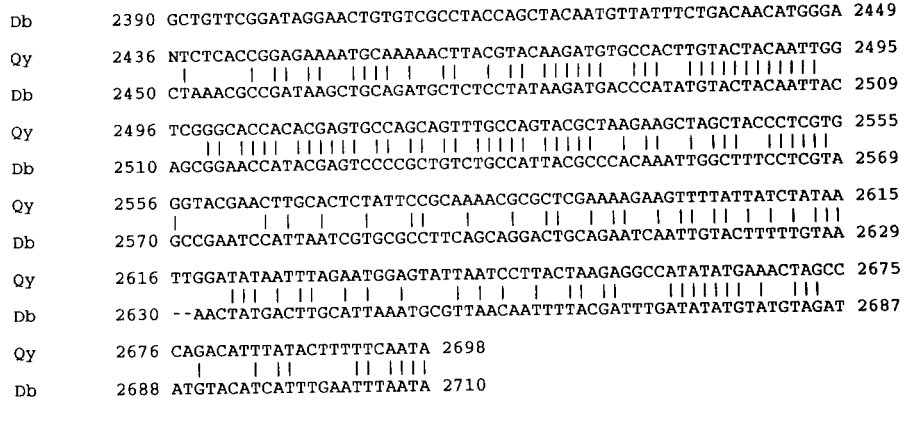
Qy      1137 GCTCTAAAAACTAACGCTAGCGAATTAGTGGTACTAATTCCTGAGCTCTGCCGAGTGACT 1196
              | |      |   | | |     |  | | ||||| |||||    || |  ||
Db      1136 AATATTCGTGGTGGAAATGACCAGGCCATAATGATCATTCCCGAGCTGGCACGGGCTACG 1195

Qy      1197 GGGCTCAATGCCGAGATGCGCTCAAACTTTCAGCTTATGCGTGCCATGAGCAGTTATACG 1256
             ||  | |  | ||  |||||| |  |||||  |    || | ||||||||     ||||
Db      1196 GGAATGACGGACGCTATGCGCGCTGACTTTAGGACGTTGAGGGCCATGAGTGAACATACC 1255

Qy      1257 CGAATGAACCCCAAACAACGCACTGATCGATTGCGCGCTTTTAACCACCGTTTACAAAAC 1316
              |  |||| ||  | |  |||| || ||  | |||   || ||| | || ||| |
Db      1256 AGGCTGAATCCAGATCGTCGCATCGAACGCCTTCGCATGTTCAACAAGCGCTTAAAGTCG 1315

Qy      1317 ACTCCAGAAAGTGTGAAGGTCTTGAGAGACTGGAACATGGAACTGGACAAGAACGTCACA 1376
```





RESULT 8
AY014899
LOCUS AY014899 3057 bp mRNA linear INV 02-JAN-2001
DEFINITION Strongylocentrotus purpuratus seawi mRNA, complete cds.
ACCESSION AY014899
VERSION AY014899.1 GI:12007639
KEYWORDS .
SOURCE Strongylocentrotus purpuratus.
ORGANISM Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinoidea; Echinacea; Echinoida;
Strongylocentrotidae; Strongylocentrotus.
REFERENCE 1 (bases 1 to 3057)
AUTHORS Rodriguez,A.J. and Bonder,E.M.
TITLE Seawi - Cloning and Molecular Characterization of a Sea Urchin Homolog of Piwi
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3057)
AUTHORS Rodriguez,A.J. and Bonder,E.M.
TITLE Direct Submission
JOURNAL Submitted (28-NOV-2000) Biological Sciences, Rutgers University-Newark, 101 Warren Street Room 135, Newark, NJ 07102, USA

FEATURES Location/Qualifiers
source 1. .3057
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
CDS 181. .2745
/codon_start=1
/product="seawi"
/protein_id="AAG42533.1"
/db_xref="GI:12007640"
/translation="MDRRPGEQAPPPMQQAAPPQEASSGMPVPGRGRSRGGTTQPGPG
AVPKQEPRGELTSQIAKLAIGASGEGRVGGAVQDRRRRRDEIELEPTTRPENFVKQAI
AGDKIALIANGFKLKTKPDWQLYQYRVDFEPEILNPRARFALLKGHSALLGKGLTLDM
DTMYSLFKLAEKVTNLSAERKDGSNVNICVTHVATLNPMAPNTLHLYNVLFRRCLKMI
NMEQVGRNYYDPTAAIDIKQHGLQLWPGFVTSILQYEYDVMLLSDISHKVLRTQTVWE
VMNDLFNKARGRFKEEITKLMIGQIVLTKYNNKTYRVDDIDFDTTPADTFETRSGPVS
YVDYFKKSYERVIHDVNQPMLVSRPKKREEKKGVGPAYLPPELCFLTGLSDDMRADFN
VMKDMGQHTRVGPQDRCRTLSGFIKKLSSNEEVKTYLDSWGMEFDKEQVKLTGRVLPP
EKLFQRGKQFSYNPSNADWSRDTRGNALTDAKILNNWKIFYTRRDANRGQDFIKSLVR
VANPMGMNVRGPEIVELPDDRTETYTRSLQAQIAQDTQIVVVILPTNRKDRYDAIKKT
CVVTHPCPSQVIVSRTLSKQQMLMSVATKIAMQMNCKMGGDLWRVEIPLSNLMIIGID
SYHDSLTKGRSVLGFVASMNKSQTSFFSSCAFQHAQGEFGANLSTLMNNALKRYYQIN
EKFPERIIIFRDGVGDSQVNLVVDYELKQIKDTLDKVYPQGTVHKLAVVVVKKRINNR
FFANLRGGLSNPPPGTVIDDVVTKPHLYDYFIISQSVRQGSVSPTSYNVVYDTTGLKP
DHMQRLTYKLTHLYFNWPGTVRVPAPCMYAHKLAFLIGQSVHEVPNNRLADTLFYL"

BASE COUNT 822 a 806 c 796 g 633 t
ORIGIN

Query Match 10.6%; Score 321.8; DB 3; Length 3057;
Best Local Similarity 48.0%; Pred. No. 4.2e-59;
Matches 1167; Conservative 0; Mismatches 1241; Indels 23; Gaps 8;

```
Qy       201 GACCCGAGAGCGGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGAAGCT 260
Db       328 GAACCTAGAGGGGAGCTCACGAGTCAGATTGCCAAGTTGGCCATTGGAGCCAGCGGTGAA 387

Qy       261 CCCAGGCGTGAAGGTGGCCCGCCAGAGCGAAAGCCGTGGGGTGACCAATATGATTACCTG 320
Db       388 GGAAGAGTCGGGGGTGCGGTGCAGGATAGGCGCAGGAGAAGGGACGAGATTGAACTAGAG 447

Qy       321 AATACCCGTCCGGTTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTG 380
Db       448 CCGACCACTCGACCTGAGAACTTTGTCAAGCAGGCAATTGCTGGAGATAAGATCGCACTG 507

Qy       381 CAGACGAACTTTTTTCGANTAAAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTG 440
Db       508 ATAGCAAACGGCTTCAAGTTGAAGACTAAACCAGACTGGCAACTCTACCAGTACCGGGTT 567

Qy       441 GAGTTTGTGCCGACCATCGAGAATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCT 500
Db       568 GACTTTGAGCCTGAGATCCTGAACCCTAGAGCTCGTTTTGCCCTCCTGAAGGGCCACAGT 627

Qy       501 AACCTTCTGGGATCAGGCTATCTATTCGACGGACTGCAACTGTTCACCACCAGGAAATTC 560
Db       628 GCTCTTCTTGGTAAAGGGTTAACGCTCGATATGGACACCATGTATTCTCTTTTCAAGCTA 687

Qy       561 GAGCAGGAAATCACGGTGCTCAGCGGAAAGTCGAAGCTGGACATTGAATACAAGATATCC 620
Db       688 GCAGAAAAGGTAACCAACCTGAGCGCAGAGAGGAAGGATGGGTCTAACGTGAATATCTGT 747

Qy       621 ATAAAGTTCGTTGGATTCATATCGTGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTA 680
Db       748 GTGACTCATGTTGCTACTCTTAACCCCATGGCACCAAACACCCTGCATCTCTACAACGTC 807

Qy       681 ATATTGCGCCGCTCGATGAAGGGCCTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGAT 740
Db       808 CTCTTTAGAAGGTGTTTGAAGATGATCAATATGGAGCAAGTTGGACGCAACTACTATGAT 867

Qy       741 CCCCGAGCTAAGATCGAAATAAGGGAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACA 800
Db       868 CCTACGGCCGCCATTGACATCAAGCAGCATGGGCTGCAGCTGTGGCCAGGGTTCGTGACA 927

Qy       801 TCGATTCGTCAGCACGAAAAAGATATTTTATTGGGCACCGAAATAACTCACAAAGTTATG 860
Db       928 TCCATCCTTCAGTACGAGTACGACGTGATGCTGCTGTCTGACATCAGCCACAAGGTTCTG 987

Qy       861 CGCACCGAGACGATCTACGACATAATGCGACGTTGCTCACACAA---TCCGGCTCGTCAT 917
Db       988 CGTACCCAGACCGTCTGGGAGGTCATGAATGATCTCTTCAACAAAGCTAGGGGCAGGTTC 1047

Qy       918 CAGGACGAAGTACGGGTAAATGTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGA 977
Db      1048 AAGGAGGAAATCACAAAGCTCATGATTGGACAGATTGTCCTCACAAAGTACAACAACAAG 1107

Qy       978 ACTTATCGTATCAATGATGTCGACTTTGGACAAACTCCGAAATCAACATTCAGTTGCAAG 1037
Db      1108 ACCTACCGGGTAGATGATATCGACTTTGACACTACTCCAGCGGATACTTTTGAGACCCGT 1167

Qy      1038 GGTAGAGATATCAGTTTCGTGGAATACTATCTCACTAAATATAATATACGCATTCGCGAC 1097
Db      1168 TCCGGACCAGTTTCCTATGTCGATTACTTCAAGAAGTCCTATGAGAGGGTCATTCACGAC 1227

Qy      1098 CACAATCAGCCGCTGCTGATTTCCAAAAATAGGGACAAGGCTCTAAAAACTAACGCTAGC 1157
Db      1228 GTGAACCAACCCATGCTGGTGTCTCGCCCAAAGAAGAGGG---AGGAAAAGAAGGGAGTT 1284

Qy      1158 GAATTAGTGGTACTAATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGC 1217
Db      1285 GGTCCAGCCTACCTACCGCCTGAGCTGTGCTTCCTGACTGGCCTAAGCGACGACATGCGC 1344

Qy      1218 TCAAACTTTCAGCTTATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGC 1277
```

```
Db      1345 GCCGATTTCAACGTCATGAAGGACATGGGCCAACACACTCGTGTTGGACCCCAAGACAGA 1404

Qy      1278 ACTGATCGATTGCGCGCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTC 1337
Db      1405 TGTAGAACGCTCAGCGGCTTCATCAAAAAGCTCAGCAGCAATGAAGAAGTGAAGACTTAC 1464

Qy      1338 TTGAGAGACTGGAACATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATT 1397
Db      1465 CTAGATTCCTGGGGCATGGAGTTCGACAAAGAGCAGGTCAAACTGACTGGGCGTGTCCTG 1524

Qy      1398 GGACAGCAGAACATCGTGTTTCATAATGGAAAGGTTCCTGCTGGAGAAAACGCTGATTGG 1457
Db      1525 CCACCTGAGAAACT----ATTCCAGAGGGGTAAACAGTTCAGTTACAATCCGAGCAATGC 1580

Qy      1458 CAAAGGCACTTCAGAGACCAAAGGATGCTTAC-CACTCCGAGCGATGGCCTCGATCGTTG 1516
Db      1581 GGACTGGTCCAGAGACACCAGGGGTAATGCCCTCACTGATGCAAAGATCCTCAACAACTG 1640

Qy      1517 GGCTGTCATCGCGCCGCAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTA 1576
Db      1641 GAAGATCTTCTACACACGCCGGGACGCCAACAGGGGCCAGGACTTCATCAAGTCACTCGT 1700

Qy      1577 TAGAGCAGCTAGTGGAATGGGTCTTAGAATTCGAAGCCCCCAGGAATTCATAATTTATGA 1636
Db      1701 ACGAGTCGCAAACCCAATGGGTATGAACGTACGAGGACCAGAGATAGTGGAACTTCCCGA 1760

Qy      1637 TGATCGCACTGGAACTTATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACT 1696
Db      1761 TGATCGTACAGAGACCTACACAAGATCCCTACAGGCACAAATCGCACAGGATACACAGAT 1820

Qy      1697 TATATTATGCCTCGTACCCAATGATAACGCCGAAAGATACTCATCAATCAAAAAGAGAGG 1756
Db      1821 TGTGGTGGTGATCCTGCCAACGAACAGAAAGGACCGGTACGATGCCATCAAGAAGACGTG 1880

Qy      1757 ATACGTTGACAGGGCGGTGCCAACTCAAG---TTGTGACCCTTAAAACGACCAAGAACCG 1813
Db      1881 CGTGGTGACCCATCCATGCCCCAGCCAGGTCATCGTCTCCAGGACCCTGTCCAAGCAGCA 1940

Qy      1814 TAGCCTTATGAGCATTGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATAC 1873
Db      1941 GATGCTCATGTCTGTAGCCACCAAGATTGCCATGCAGATGAACTGCAAGATGGGCGGAGA 2000

Qy      1874 ACCCTGGATGATCGAACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAA 1933
Db      2001 CCTCTGGCGAGTCGAGATCCCGCTGAGCAACCTGATGATCATCGGTATTGACAGCTACCA 2060

Qy      1934 GAGCACACGAGATCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCA 1993
Db      2061 TGACTCTCTGACCAAAGGCCGCTCCGTCCTGGGATTCGTGGCTTCCATGAACAAGTCTCA 2120

Qy      1994 AAACTCCACGTACTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACAC 2053
Db      2121 AA---CCAGCTTCTTCTCTAGTTGTGCATTCCAGCACGCACAGGGAGAATTCGGTGCTAA 2177

Qy      2054 CCTTTGGCCGATGATAGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCC 2113
Db      2178 CCTGAGTACCCTCATGAACAATGCCCTGAAGCGATACTATCAGATCAACGAGAAGTTTCC 2237

Qy      2114 ATCTCGAATCGTATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGA 2173
Db      2238 TGAGCGTATCATCATCTTCCGTGATGGTGTTGGTGACAGCCAAGTCAATCTGGTCGTTGA 2297

Qy      2174 ATTTGAAGTCAAGGACATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCT--- 2230
Db      2298 CTATGAGCTCAAACAAATCAAGGACACCTTGGACAAAGTCTACCCTCAAGGGACGGTTCA 2357

Qy      2231 -AAGCCCACCGCAATTAGCTTATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCC 2289
Db      2358 CAAGCTGGCCGTTGTGGTCGTCAAGAAGCGCATCAATAACCGCTTCTTTGCCAATCTGAG 2417

Qy      2290 TCAA--CGGACAAAATCCTCCGCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGA 2347
```

```
Db       2418 AGGAGGCCTGTCCAACCCTCCCCCTGGAACCGTCATCGATGATGTCGTCACCAAGCCCCA 2477

Qy       2348 GAGATACGACTTTTATCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAG 2407

Db       2478 CCTCTATGACTACTTCATCATCAGTCAGAGTGTCCGCCAGGGCTCCGTGTCCCCCACGTC 2537

Qy       2408 CTACAATGTTCTTTATAGCAGCATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTA 2467

Db       2538 CTACAACGTTGTCTATGACACCACTGGTCTCAAGCCAGACCACATGCAGAGGTTGACCTA 2597

Qy       2468 CAAGATGTGCCACTTGTACTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCA 2527

Db       2598 CAAGCTGACCCATCTCTACTTCAACTGGCCGGGAACGGTGCGTGTTCCTGCCCCGTGCAT 2657

Qy       2528 GTACGCTAAGAAGCTAGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGC 2587

Db       2658 GTATGCACACAAGCTGGCATTCCTGATCGGGCAGAGCGTCCATGAAGTGCCCAACAACAG 2717

Qy       2588 GCTCGAAAAGAAGTTTTATTATCTATAATTG 2618

Db       2718 ACTAGCCGACACCCTCTTCTATCTCTAGATG 2748


RESULT 9
AB032604
LOCUS       AB032604               3910 bp    mRNA    linear   ROD 04-OCT-2001
DEFINITION  Mus musculus mRNA for MIWI, complete cds.
ACCESSION   AB032604
VERSION     AB032604.1  GI:7416110
KEYWORDS    MIWI (piwi).
SOURCE      Mus musculus tissue_lib:testis cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Kuramochi-Miyagawa,S., Kimura,T., Yomogida,K., Kuroiwa,A.,
            Tadokoro,Y., Fujita,Y., Sato,M., Matsuda,Y. and Nakano,T.
  TITLE     Two mouse piwi-related genes: miwi and mili
  JOURNAL   Mech. Dev. 108 (1-2), 121-133 (2001)
  MEDLINE   21463379
REFERENCE   2  (bases 1 to 3910)
  AUTHORS   Miyagawa,S.K. and Nakano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-1999) Satomi K Miyagawa, Research Institute for
            Microbial Diseases, Department of Molecular Cell Biology; Yamadaoka
            3-1, Suita, Osaka 565-0871, Japan
            (E-mail:smiyagaw@biken.osaka-u.ac.jp, Tel:81-6-6879-8362,
            Fax:81-6-6879-8362)
FEATURES             Location/Qualifiers
     source          1. .3910
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /map="5F distal-G2 proximal"
                     /tissue_lib="testis"
     gene            1. .3910
                     /gene="miwi"
     CDS             156. .2744
                     /gene="miwi"
                     /note="piwi family"
                     /codon_start=1
                     /product="MIWI"
                     /protein_id="BAA93705.1"
                     /db_xref="GI:7416111"
                     /translation="MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEG
                     DLVGRGRQRGMVVGATSKSQELQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVK
                     ESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGR
                     CHAFDGTILFLPKRLQHKVTEVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIIFR
                     RLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLR
                     SETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFK
                     KADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTG
                     LTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNL
                     LSFSGRILQSEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEA
                     ANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTEAYLRALQQKVTSDTQIVVCLLSSNRK
                     DKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMPL
                     KLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQA
                     ALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIV
                     VKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVI
                     YDSSGLKPDHIQRLTYKLCHVYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLS
                     NRLYYL"
BASE COUNT      993 a    970 c   1063 g    884 t
ORIGIN

  Query Match             10.3%;  Score 314.4;  DB 10;  Length 3910;
  Best Local Similarity   48.8%;  Pred. No. 1.7e-57;
  Matches 1118;  Conservative    0;  Mismatches 1143;  Indels   29;  Gaps    9;

Qy        345 TCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGANTAAAA 404

Db        465 TCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCACTTCCGGCTGACC 524

Qy        405 ACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAGAAT 464

Db        525 TCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCCCTGATGGAGGCC 584

Qy        465 CCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTATCTA 524

Db        585 CGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGAAGGTGTCATGCT 644

Qy        525 TTCGACGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTCAGC 584

Db        645 TTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGTCACAGAAGTATTC 704

Qy        585 GGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATATCG 644

Db        705 AGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCACCCTGACCAACGAGCTGCCG 764

Qy        645 TGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCGATGAAGGGC 704

Db        765 CCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCATCTTCAGGAGGCTCTTGAAAATC 824

Qy        705 CTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATAAGG 764

Db        825 ATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAGTGACCCGATTGATATTCCA 884

Qy        765 GAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAAGAT 824

Db        885 AACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCATCCTTCAGTATGAGAACAAC 944

Qy        825 ATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGACATA 884

Db        945 ATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAGCGAGACTGTCCTAGACTTC 1004

Qy        885 ATGCGACGTTGCTCACACAATCCGGCTCGTCA------TCAGGACGAAGTACGGGTAAAT 938

Db       1005 ATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCAGGAGCAAGTGTCGAAGGAG 1064

Qy        939 GTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGATGTC 998

Db       1065 CTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGACCTACCGGGTGGATGACATT 1124

Qy        999 GACTTTGGACAAACTCCGAAATCAACATTCAGTTGCAAGGGTAG---AGATATCAGTTTC 1055

Db       1125 GACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGATGGCTCGGAGGTCAGCTTC 1184

Qy       1056 GTGGAATACTATCTCACTAAATATAATATACGCATTCGCGACCACAATCAGCCGCTGCTG 1115

Db       1185 CTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGACCTGAAGCAGCCGGTGCTG 1244

Qy       1116 ATTTCCAAA---AATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTACTA 1172

Db       1245 GTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCTGCCTGGCCCAGCTATGCTC 1304

Qy       1173 ATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCAGCTT 1232

Db       1305 ATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAATGCGCAATGATTTCAATGTG 1364
```

```
Qy       1233 ATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATCGATTGCGC 1292
Db       1365 ATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCAGCGGCAGCGGGAGGTGGGC 1424

Qy       1293 GCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGACTGGAAC 1352
Db       1425 CGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAGAGAGCTTCGAGACTGGGGC 1484

Qy       1353 ATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCAGAACATC 1412
Db       1485 CTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAATCTTACAATCTGAGAAGATC 1544

Qy       1413 GTGTTTCATAATGGAAAGGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTCAGA 1472
Db       1545 ----CACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGCAGACTGGTCCAAAGA 1600

Qy       1473 GACCAAAGGATGCTTACCACTCCGAGCGATGGC-CTCGATCGTTGGGCTGTCATCGCGCC 1531
Db       1601 AACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGGCTGCTGATCTATAC 1660

Qy       1532 GCAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAGTGG 1591
Db       1661 CCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTTCAAAGTGACTCCAGC 1720

Qy       1592 AATGGGTCTTAGAATTCGAAGCCCCCAGGAATTCATAATTTATGATGATCGCACTGGAAC 1651
Db       1721 CATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGT---GGATGACAGAACAGAAGC 1777

Qy       1652 TTATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGCCTCGT 1711
Db       1778 TTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGTCTCTT 1837

Qy       1712 ACCCAATGATAACGCCGAAAGATACTCATCAATCAAAAAGAGAGGATACGTTGACAGGGC 1771
Db       1838 GTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTGTACAGACTGCCC 1897

Qy       1772 GGTGCCAACTCA---AGTTGTGACCCTTAAAACGACCAAGAACCGTAGCCTTATGAGCAT 1828
Db       1898 CACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATGGCCAT 1957

Qy       1829 TGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGATCGA 1888
Db       1958 TGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGGGTGGA 2017

Qy       1889 ACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGATCG 1948
Db       2018 CATGCCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACACCACAGCTGG 2077

Qy       1949 GAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTT 2008
Db       2078 GCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTC 2137

Qy       2009 CAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATGAT 2068
Db       2138 CCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGCTTGCA 2197

Qy       2069 AGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATT 2128
Db       2198 AGC---TGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATCGT 2254

Qy       2129 TTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAAGGA 2188
Db       2255 GTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCACA 2314

Qy       2189 CATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATTAGC 2248
Db       2315 GTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATCGT 2374

Qy       2249 TTATATTGTGGTAACCAGATCCATGAA---CACGCGCTTCTTCCTCAACGGACAAAATCC 2305
Db       2375 GGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAACCC 2434

Qy       2306 TCCGCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTATCT 2365

Db       2435 TCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTCAT 2494

Qy       2366 GGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAG 2425
Db       2495 CGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTATGA 2554

Qy       2426 CAGCATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTA 2485
Db       2555 CAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGCTCTGCCACGTGTA 2614

Qy       2486 CTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGC 2545
Db       2615 CTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTGGC 2674

Qy       2546 TACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTA 2605
Db       2675 CTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTCTA 2734

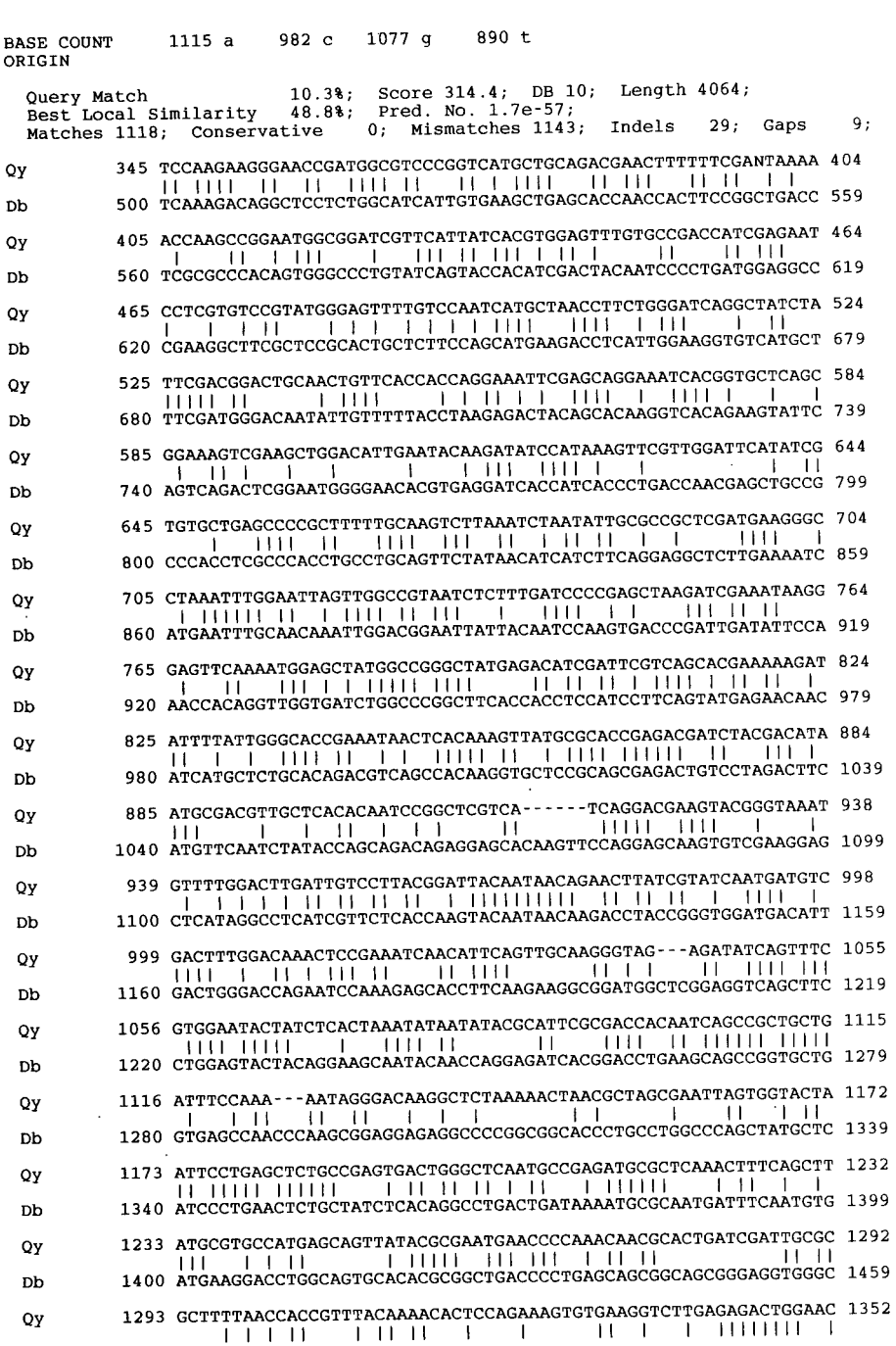
Qy       2606 TTATCTATAA 2615
Db       2735 CTACCTCTAA 2744
```

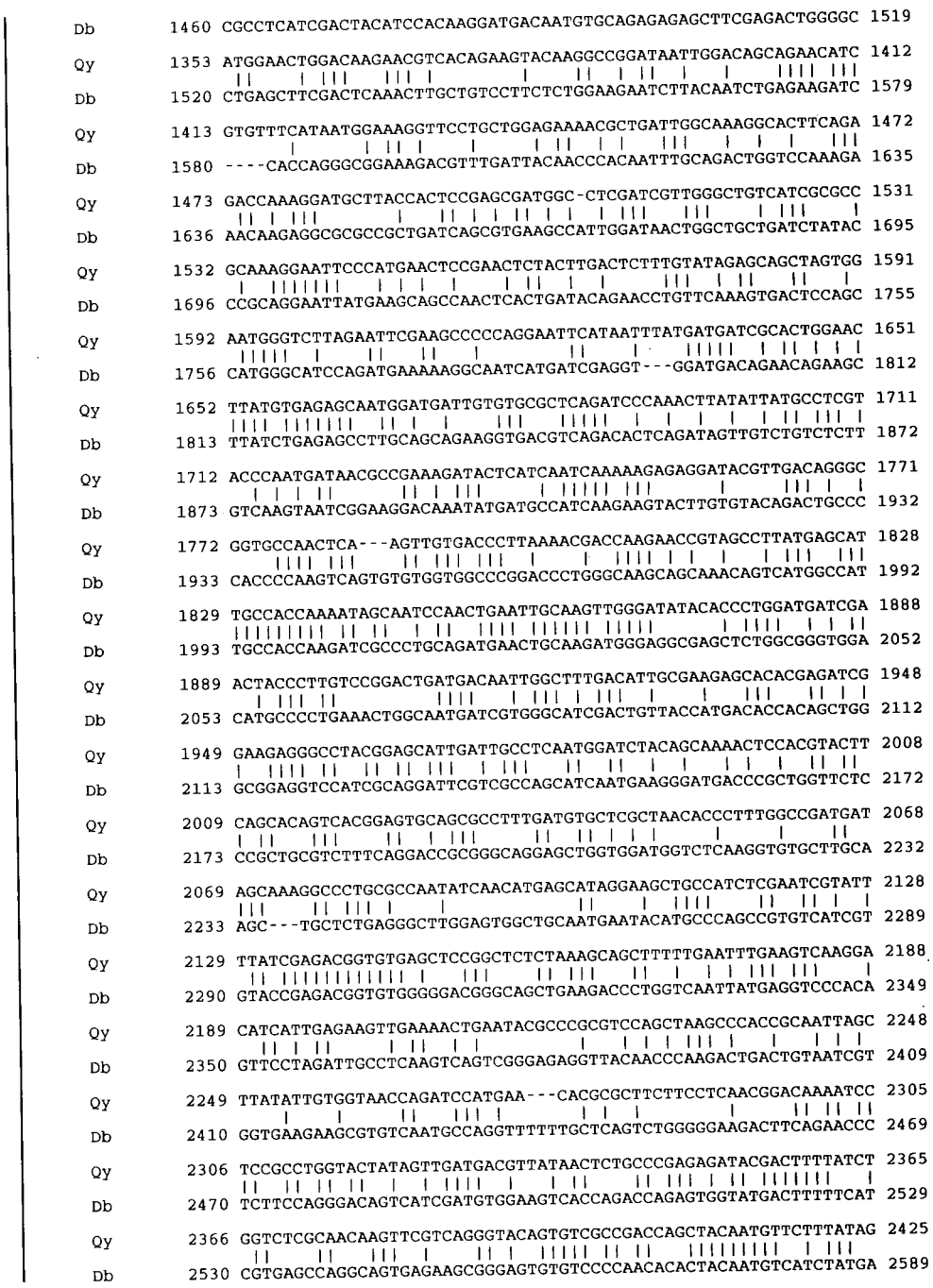
RESULT 10
AF438405

```
LOCUS       AF438405                4064 bp    mRNA    linear   ROD 13-NOV-2001
DEFINITION  Mus musculus MIWI mRNA, complete cds.
ACCESSION   AF438405
VERSION     AF438405.1  GI:16905060
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4064)
  AUTHORS   Deng,W. and Lin,H.
  TITLE     Miwi, a murine homolog of piwi, encodes a cytoplasmic protein
            essential for spermatogenesis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4064)
  AUTHORS   Deng,W. and Lin,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2001) Department of Cell Biology, Duke
            University, Room 412, Nanaline Duke Building, Durham, NC 27710, USA
FEATURES             Location/Qualifiers
     source          1. .4064
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /tissue_type="testis"
     CDS             191. .2779
                     /note="Piwi-like protein"
                     /codon_start=1
                     /product="MIWI"
                     /protein_id="AAL31014.1"
                     /db_xref="GI:16905061"
                     /translation="MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEG
                     DLVGRGRQRGMVVGATSKSQELQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVK
                     ESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGR
                     CHAFDGTILFLPKRLQHKVTEVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIIFR
                     RLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLR
                     SETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFK
                     KADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTG
                     LTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNL
                     LSFSGRILQSEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEA
                     ANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTEAYLRALQQKVTSDTQIVVCLLSSNRK
                     DKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMPL
                     KLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQA
                     ALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIV
                     VKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVI
                     YDSSGLKPDHIQRLTYKLCHVYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLS
                     NRLYYL"
```

BASE COUNT 1115 a 982 c 1077 g 890 t
ORIGIN

Query Match 10.3%; Score 314.4; DB 10; Length 4064;
Best Local Similarity 48.8%; Pred. No. 1.7e-57;
Matches 1118; Conservative 0; Mismatches 1143; Indels 29; Gaps 9;





```
Qy       2426 CAGCATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTA 2485
Db       2590 CAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGCTCTGCCACGTGTA 2649

Qy       2486 CTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGC 2545
Db       2650 CTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTGGC 2709

Qy       2546 TACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTA 2605
Db       2710 CTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTCTA 2769

Qy       2606 TTATCTATAA 2615
Db       2770 CTACCTCTAA 2779
```

```
RESULT 11
AY014900
LOCUS       AY014900               2605 bp    mRNA    linear   INV 02-JAN-2001
DEFINITION  Strongylocentrotus purpuratus clone 6-2 seawi mRNA, partial cds.
ACCESSION   AY014900
VERSION     AY014900.1  GI:12007641
KEYWORDS    .
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 2605)
  AUTHORS   Rodriguez,A.J. and Bonder,E.M.
  TITLE     Seawi - Cloning and Molecular Characterization of a Sea Urchin
            Homolog of Piwi
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2605)
  AUTHORS   Rodriguez,A.J. and Bonder,E.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-2000) Biological Sciences, Rutgers
            University-Newark, 101 Warren Street Room 135, Newark, NJ 07102,
            USA
FEATURES             Location/Qualifiers
     source          1. .2605
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="6-2"
     CDS             181. .>2605
                     /codon_start=1
                     /product="seawi"
                     /protein_id="AAG42534.1"
                     /db_xref="GI:12007642"
                     /translation="MDRRPGEQAPPPMQQAAPPQEASSGMPVPGRGRSRGGTTQPGPG
                     AVPKQEPRGELTSQIAKLAIGASGEGRVGGAVQDRRRRRDEIELEPTTRPENFVKQAI
                     AGDKIALIANGFKLKTKPDWQLYQYRVDFEPEILNPRARFALLKGHSALLGKGLTLDM
                     DTMYSLFKLAEKVTNLSAERKDGSNVNICVTHVATLNPMAPNTLHLYNVLFRRCLKMI
                     NMEQVGRNYYDPTAAIDIKQHGLQLWPGFVTSILQYEYDVMLLSDISHKVLRTQTVWE
                     VMNDLFNKARGRFKEEITKLMIGQIVLTKYNNKTYRVDDIDFDTTPADTFETRSGPVS
                     YVDYFKKSYERVIHDVNQPMLVSRPKKREEKKGVGPAYLPPELCFLTGLSDDMRADFN
                     VMKDMGQHTRVGPQDRCRTLSGFIKKLSSNEEVKTYLDSWGMEFDKEQVKLTGRVLPP
                     EKLFQRGKQFSYNPSNADWSRDTRGNALTDAKILNNWKIFYTRRDANRGQDFIKSLVR
                     VANPMGMNVRGPEIVELPDDRTETYTRSLQAQIAQDTQIVVVILPTNRKDRYDAIKKT
                     CVVTHPCPSQVIVSRTLSKQQMLMSVATKIAMQMNCKMGGDLWRVEIPLSNLMIIGID
                     SYHDSLTKGRSVLGFVASMNKSQTSFFSSCAFQHAQGEFGANLSTLMNNALKRYYQIN
                     EKFPERIIIFRDGVGDSQVNLVVDYELKQIKDTLDKVYPQGTVHKLAVVVVKKRINNR
                     FFANLRGGLSNPPPGTVIDDVVTKPHLYDYFIISQSVRQGSVSPTSYNVVYDTTGLKP
                     DHMQRLTYGI"
BASE COUNT      718 a    695 c    686 g    506 t
ORIGIN
```

```
Query Match             9.1%;  Score 277.6;  DB 3;  Length 2605;
Best Local Similarity   47.5%;  Pred. No. 1.5e-49;
Matches 1085;  Conservative    0;  Mismatches 1178;  Indels   23;  Gaps    8;
```

```
Qy        201 GACCCGAGAGCGGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGAAGCT 260
Db        328 GAACCTAGAGGGGAGCTCACGAGTCAGATTGCCAAGTTGGCCATTGGAGCCAGCGGTGAA 387

Qy        261 CCCAGGCGTGAAGGTGGCCCGCCAGAGCGAAAGCCGTGGGGTGACCAATATGATTACCTG 320
Db        388 GGAAGAGTCGGGGGTGCGGTGCAGGATAGGCGCAGGAGAAGGGACGAGATTGAACTAGAG 447

Qy        321 AATACCCGTCCGGTTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTG 380
Db        448 CCGACCACTCGACCTGAGAACTTTGTCAAGCAGGCAATTGCTGGAGATAAGATCGCACTG 507

Qy        381 CAGACGAACTTTTTTCGANTAAAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTG 440
Db        508 ATAGCAAACGGCTTCAAGTTGAAGACTAAACCAGACTGGCAACTCTACCAGTACCGGGTT 567

Qy        441 GAGTTTGTGCCGACCATCGAGAATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCT 500
Db        568 GACTTTGAGCCTGAGATCCTGAACCCTAGAGCTCGTTTTGCCCTCCTGAAGGGCCACAGT 627

Qy        501 AACCTTCTGGGATCAGGCTATCTATTCGACGGACTGCAACTGTTCACCACCAGGAAATTC 560
Db        628 GCTCTTCTTGGTAAAGGGTTAACGCTCGATATGGACACCATGTATTCTCTTTTCAAGCTA 687

Qy        561 GAGCAGGAAATCACGGTGCTCAGCGGAAAGTCGAAGCTGGACATTGAATACAAGATATCC 620
Db        688 GCAGAAAAGGTAACCAACCTGAGCGCAGAGAGGAAGGATGGGTCTAACGTGAATATCTGT 747

Qy        621 ATAAAGTTCGTTGGATTCATATCGTGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTA 680
Db        748 GTGACTCATGTTGCTACTCTTAACCCCATGGCACCAAACACCCTGCATCTCTACAACGTC 807

Qy        681 ATATTGCGCCGCTCGATGAAGGGCCTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGAT 740
Db        808 CTCTTTAGAAGGTGTTTGAAGATGATCAATATGGAGCAAGTTGGACGCAACTACTATGAT 867

Qy        741 CCCCGAGCTAAGATCGAAATAAGGGAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACA 800
Db        868 CCTACGGCCGCCATTGACATCAAGCAGCATGGGCTGCAGCTGTGGCCAGGGTTCGTGACA 927

Qy        801 TCGATTCGTCAGCACGAAAAAGATATTTTATTGGGCACCGAAATAACTCACAAAGTTATG 860
Db        928 TCCATCCTTCAGTACGAGTACGACGTGATGCTGCTGTCTGACATCAGCCACAAGGTTCTG 987

Qy        861 CGCACCGAGACGATCTACGACATAATGCGACGTTGCTCACACAA---TCCGGCTCGTCAT 917
Db        988 CGTACCCAGACCGTCTGGGAGGTCATGAATGATCTCTTCAACAAAGCTAGGGGCAGGTTC 1047

Qy        918 CAGGACGAAGTACGGGTAAATGTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGA 977
Db       1048 AAGGAGGAAATCACAAAGCTCATGATTGGACAGATTGTCCTCACAAAGTACAACAACAAG 1107

Qy        978 ACTTATCGTATCAATGATGTCGACTTTGGACAAACTCCGAAATCAACATTCAGTTGCAAG 1037
Db       1108 ACCTACCGGGTAGATGATATCGACTTTGACACTACTCCAGCGGATACTTTTGAGACCCGT 1167

Qy       1038 GGTAGAGATATCAGTTTCGTGGAATACTATCTCACTAAATATAATATACGCATTCGCGAC 1097
Db       1168 TCCGGACCAGTTTCCTATGTCGATTACTTCAAGAAGTCCTATGAGAGGGTCATTCACGAC 1227

Qy       1098 CACAATCAGCCGCTGCTGATTTCCAAAAATAGGGACAAGGCTCTAAAAACTAACGCTAGC 1157
Db       1228 GTGAACCAACCCATGCTGGTGTCTCGCCCAAAGAAGAGGG---AGGAAAAGAAGGGAGTT 1284

Qy       1158 GAATTAGTGGTACTAATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGC 1217
Db       1285 GGTCCAGCCTACCTACCGCCTGAGCTGTGCTTCCTGACTGGCCTAAGCGACGACATGCGC 1344

Qy       1218 TCAAACTTTCAGCTTATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGC 1277
Db       1345 GCCGATTTCAACGTCATGAAGGACATGGGCCAACACACTCGTGTTGGACCCCAAGACAGA 1404

Qy       1278 ACTGATCGATTGCGCGCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTC 1337
```

```
Db       1405 TGTAGAACGCTCAGCGGCTTCATCAAAAAGCTCAGCAGCAATGAAGAAGTGAAGACTTAC 1464

Qy       1338 TTGAGAGACTGGAACATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATT 1397

Db       1465 CTAGATTCCTGGGGCATGGAGTTCGACAAAGAGCAGGTCAAACTGACTGGGCGTGTCCTG 1524

Qy       1398 GGACAGCAGAACATCGTGTTTCATAATGGAAAGGTTCCTGCTGGAGAAAACGCTGATTGG 1457

Db       1525 CCACCTGAGAAACT----ATTCCAGAGGGGTAAACAGTTCAGTTACAATCCGAGCAATGC 1580

Qy       1458 CAAAGGCACTTCAGAGACCAAAGGATGCTTAC-CACTCCGAGCGATGGCCTCGATCGTTG 1516

Db       1581 GGACTGGTCCAGAGACACCAGGGGTAATGCCCTCACTGATGCAAAGATCCTCAACAACTG 1640

Qy       1517 GGCTGTCATCGCGCCGCAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTA 1576

Db       1641 GAAGATCTTCTACACACGCCGGGACGCCAACAGGGGCCAGGACTTCATCAAGTCACTCGT 1700

Qy       1577 TAGAGCAGCTAGTGGAATGGGTCTTAGAATTCGAAGCCCCCAGGAATTCATAATTTATGA 1636

Db       1701 ACGAGTCGCAAACCCAATGGGTATGAACGTACGAGGACCAGAGATAGTGGAACTTCCCGA 1760

Qy       1637 TGATCGCACTGGAACTTATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACT 1696

Db       1761 TGATCGTACAGAGACCTACACAAGATCCCTACAGGCACAAATCGCACAGGATACACAGAT 1820

Qy       1697 TATATTATGCCTCGTACCCAATGATAACGCCGAAAGATACTCATCAATCAAAAAGAGAGG 1756

Db       1821 TGTGGTGGTGATCCTGCCAACGAACAGAAAGGACCGGTACGATGCCATCAAGAAGACGTG 1880

Qy       1757 ATACGTTGACAGGGCGGTGCCAACTCAAG---TTGTGACCCTTAAAACGACCAAGAACCG 1813

Db       1881 CGTGGTGACCCATCCATGCCCCAGCCAGGTCATCGTCTCCAGGACCCTGTCCAAGCAGCA 1940

Qy       1814 TAGCCTTATGAGCATTGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATAC 1873

Db       1941 GATGCTCATGTCTGTAGCCACCAAGATTGCCATGCAGATGAACTGCAAGATGGGCGGAGA 2000

Qy       1874 ACCCTGGATGATCGAACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAA 1933

Db       2001 CCTCTGGCGAGTCGAGATCCCGCTGAGCAACCTGATGATCATCGGTATTGACAGCTACCA 2060

Qy       1934 GAGCACACGAGATCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCA 1993

Db       2061 TGACTCTCTGACCAAAGGCCGCTCCGTCCTGGGATTCGTGGCTTCCATGAACAAGTCTCA 2120

Qy       1994 AAACTCCACGTACTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACAC 2053

Db       2121 AA---CCAGCTTCTTCTCTAGTTGTGCATTCCAGCACGCACAGGGAGAATTCGGTGCTAA 2177

Qy       2054 CCTTTGGCCGATGATAGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCC 2113

Db       2178 CCTGAGTACCCTCATGAACAATGCCCTGAAGCGATACTATCAGATCAACGAGAAGTTTCC 2237

Qy       2114 ATCTCGAATCGTATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGA 2173

Db       2238 TGAGCGTATCATCATCTTCCGTGATGGTGTTGGTGACAGCCAAGTCAATCTGGTCGTTGA 2297

Qy       2174 ATTTGAAGTCAAGGACATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCT--- 2230

Db       2298 CTATGAGCTCAAACAAATCAAGGACACCTTGGACAAAGTCTACCCTCAAGGGACGGTTCA 2357

Qy       2231 -AAGCCCACCGCAATTAGCTTATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCC 2289

Db       2358 CAAGCTGGCCGTTGTGGTCGTCAAGAAGCGCATCAATAACCGCTTCTTTGCCAATCTGAG 2417

Qy       2290 TCAA--CGGACAAAATCCTCCGCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGA 2347

Db       2418 AGGAGGCCTGTCCAACCCTCCCCCTGGAACCGTCATCGATGATGTCGTCACCAAGCCCCA 2477

Qy       2348 GAGATACGACTTTTATCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAG 2407

Db       2478 CCTCTATGACTACTTCATCATCAGTCAGAGTGTCCGCCAGGGCTCCGTGTCCCCCACGTC 2537

Qy       2408 CTACAATGTTCTTTATAGCAGCATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTA 2467

Db       2538 CTACAACGTTGTCTATGACACCACTGGTCTCAAGCCAGACCACATGCAGAGGTTGACCTA 2597

Qy       2468 CAAGAT 2473

Db       2598 CGGAAT 2603
```

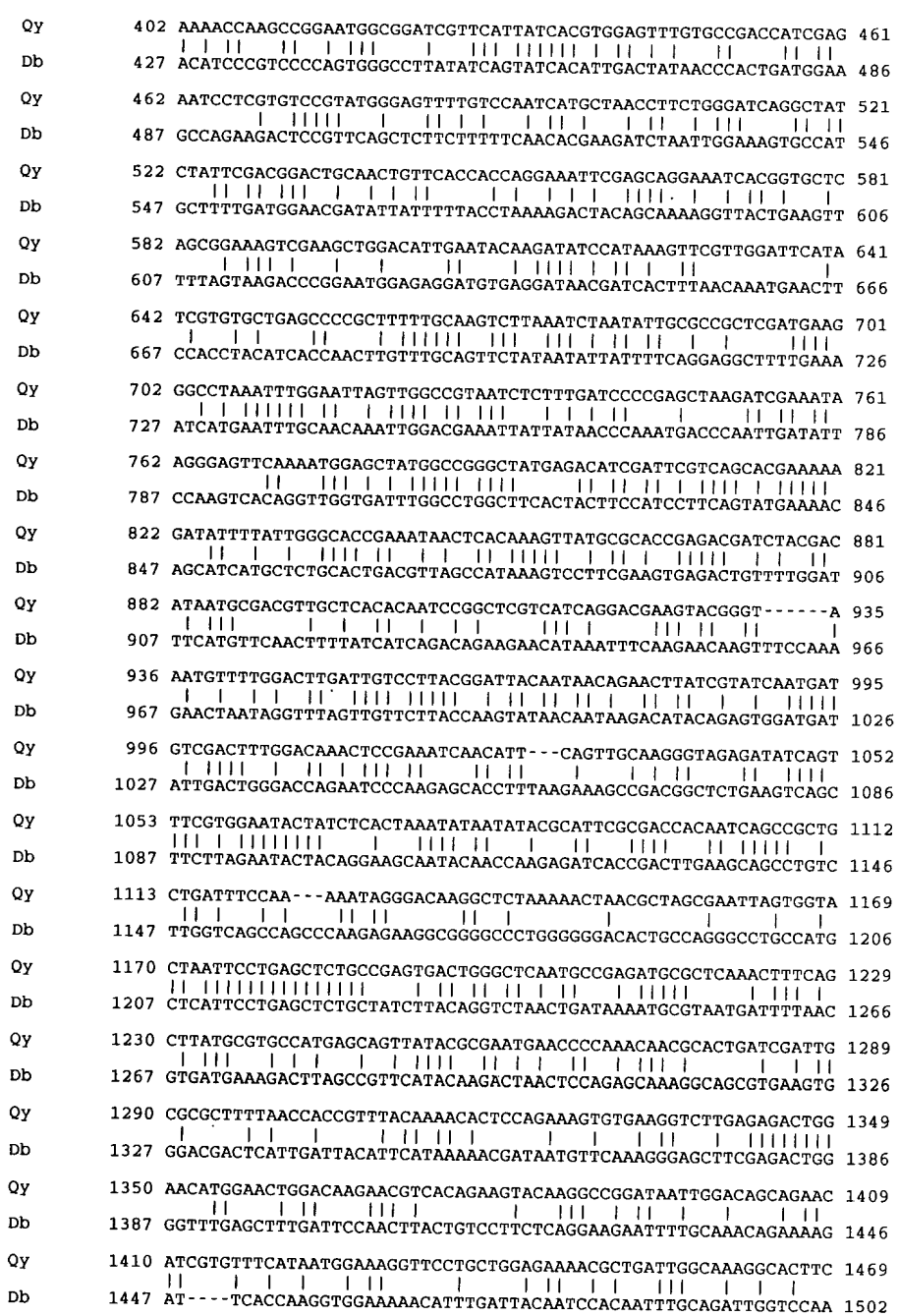
RESULT 12
AF387507

```
LOCUS       AF387507               3421 bp    mRNA    linear   PRI 01-JUL-2001
DEFINITION  Homo sapiens PIWI protein mRNA, complete cds.
ACCESSION   AF387507
VERSION     AF387507.1  GI:14579644
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3421)
  AUTHORS   Sha,J.H.
  TITLE     Cloning and identification of human piwi protein related to testis
            development
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3421)
  AUTHORS   Sha,J.H., Li,J.M. and Zhou,Z.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-2001) Key Lab of Reproductive Medicine, Nanjing
            Medical University, 140 Han Zhong Road, Nanjing Medical University,
            Jiangsu 210029, China
FEATURES             Location/Qualifiers
     source          1. .3421
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="testis"
                     /dev_stage="adult"
     CDS             64. .2649
                     /note="higher expression in adult testis than 6 month old
                     embryo testis"
                     /codon_start=1
                     /product="PIWI protein"
                     /protein_id="AAK69348.1"
                     /db_xref="GI:14579645"
                     /translation="MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEG
                     ELFGRGRQRGTAGGTAKSQGLQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVKE
                     SKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKC
                     HAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRR
                     LLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRS
                     ETVLDFMFNFYHQTEEHKFQEQVSKELIGLVVLTKYNNKTYRVDDIDWDQNPKSTFKK
                     ADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGL
                     TDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLL
                     SFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA
                     NSLIQNLFKVTPAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKD
                     KYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLK
                     LVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAA
                     LRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVV
                     KKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIY
                     DNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSN
                     RLYYL"
     polyA_signal    3354. .3359
BASE COUNT     1071 a    661 c    767 g    922 t
ORIGIN
```

```
  Query Match             8.5%;  Score 258.2;  DB 9;  Length 3421;
  Best Local Similarity   47.3%;  Pred. No. 2.5e-45;
  Matches 1084;  Conservative    0;  Mismatches 1180;  Indels   29;  Gaps    9;

Qy         342 GTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGANTA 401

Db         367 GAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTG 426
```


Qy 402 AAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAG 461
Db 427 ACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGGAA 486
Qy 462 AATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTAT 521
Db 487 GCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGCCAT 546
Qy 522 CTATTCGACGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTC 581
Db 547 GCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTT 606
Qy 582 AGCGGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATA 641
Db 607 TTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATGAACTT 666
Qy 642 TCGTGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCGATGAAG 701
Db 667 CCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAA 726
Qy 702 GGCCTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATA 761
Db 727 ATCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCAATTGATATT 786
Qy 762 AGGGAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAA 821
Db 787 CCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAAC 846
Qy 822 GATATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGAC 881
Db 847 AGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGAT 906
Qy 882 ATAATGCGACGTTGCTCACACAATCCGGCTCGTCATCAGGACGAAGTACGGGT------A 935
Db 907 TTCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAA 966
Qy 936 AATGTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGAT 995
Db 967 GAACTAATAGGTTTAGTTGTTCTTACCAAGTATAACAATAAGACATACAGAGTGGATGAT 1026
Qy 996 GTCGACTTTGGACAAACTCCGAAATCAACATT---CAGTTGCAAGGGTAGAGATATCAGT 1052
Db 1027 ATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGAAGTCAGC 1086
Qy 1053 TTCGTGGAATACTATCTCACTAAATATAATATACGCATTCGCGACCACAATCAGCCGCTG 1112
Db 1087 TTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTC 1146
Qy 1113 CTGATTTCCAA---AAATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTA 1169
Db 1147 TTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATG 1206
Qy 1170 CTAATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCAG 1229
Db 1207 CTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAAC 1266
Qy 1230 CTTATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATCGATTG 1289
Db 1267 GTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTG 1326
Qy 1290 CGCGCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGACTGG 1349
Db 1327 GGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGG 1386
Qy 1350 AACATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCAGAAC 1409
Db 1387 GGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAG 1446
Qy 1410 ATCGTGTTTCATAATGGAAAGGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTC 1469
Db 1447 AT----TCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCCAA 1502
Qy 1470 AGAGACCAAAGGATGCTTACCACTCCGAGCGATGGC-CTCGATCGTTGGGCTGTCATCGC 1528
Db 1503 AGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTA 1562
Qy 1529 GCCGCAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAG 1588
Db 1563 TACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACC 1622
Qy 1589 TGGAATGGGTCTTAGAATTCGAAGCCCCCAGGAATTCATAATTTATGATGATCGCACTGG 1648
Db 1623 AGCCATGGGCATGCAAATGAGAAAAGC---AATAATGATTGAAGTGGATGACAGAACTGA 1679
Qy 1649 AACTTATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGCCT 1708
Db 1680 AGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCT 1739
Qy 1709 CGTACCCAATGATAACGCCGAAAGATACTCATCAATCAAAAAGAGAGGATACGTTGACAG 1768
Db 1740 GTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTG 1799
Qy 1769 GGCGGTGCCAACTCA---AGTTGTGACCCTTAAAACGACCAAGAACCGTAGCCTTATGAG 1825
Db 1800 CCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGC 1859
Qy 1826 CATTGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGAT 1885
Db 1860 CATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGT 1919
Qy 1886 CGAACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGA 1945
Db 1920 GGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGC 1979
Qy 1946 TCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTA 2005
Db 1980 TGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTT 2039
Qy 2006 CTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGAT 2065
Db 2040 CTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCT 2099
Qy 2066 GATAGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGT 2125
Db 2100 GCAAGC---GGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCAT 2156
Qy 2126 ATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAA 2185
Db 2157 CGTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCC 2216
Qy 2186 GGACATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATT 2245
Db 2217 ACAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGACTAACGGTAAT 2276
Qy 2246 AGCT---TATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAA 2302
Db 2277 TGTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAA 2336
Qy 2303 TCCTCCGCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTA 2362
Db 2337 TCCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTT 2396
Qy 2363 TCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTA 2422
Db 2397 TATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTA 2456
Qy 2423 TAGCAGCATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTT 2482
Db 2457 TGACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACAT 2516
Qy 2483 GTACTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCT 2542
Db 2517 CTATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCT 2576
Qy 2543 AGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTT 2602

```
Db      2577 GGCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCT 2636

Qy      2603 TTATTATCTATAA 2615
Db      2637 TTACTACCTCTAA 2649


RESULT 13
AF264004
LOCUS       AF264004                2328 bp    mRNA    linear   PRI 21-AUG-2001
DEFINITION  Homo sapiens HIWI mRNA, complete cds.
ACCESSION   AF264004
VERSION     AF264004.1  GI:15216446
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2328)
  AUTHORS   Sharma,A.K., Nelson,M.C., Brandt,J.E., Wessman,M., Muhmud,N.,
            Weller,K.P. and Hoffman,R.
  TITLE     Human CD34+ Stem Cells Express the hiwi Gene, a Human Homolog of
            the Drosophila Gene piwi
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2328)
  AUTHORS   Sharma,A.K. and Hoffman,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-2000) Medicine, UIC, 900 South Ashland Avenue,
            Chicago, IL 60607, USA
FEATURES             Location/Qualifiers
     source          1. .2328
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="testis"
     CDS             1. .2328
                     /codon_start=1
                     /product="HIWI"
                     /protein_id="AAK92281.1"
                     /db_xref="GI:15216447"
                     /translation="MIFGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALY
                     QYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTEVFSKTRNG
                     EDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRL
                     VIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKELI
                     GLVVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVL
                     VSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQRE
                     VGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADW
                     SKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDD
                     RTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQ
                     QTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASIN
                     EGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLK
                     TLVNYEVPQFLDCLKSIGRGYNPRLTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVE
                     VTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVI
                     RVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL"
BASE COUNT      716 a    496 c    520 g    596 t
ORIGIN

  Query Match             8.4%;  Score 256.6;  DB 9;  Length 2328;
  Best Local Similarity   47.2%;  Pred. No. 5.4e-45;
  Matches 1083;  Conservative    0;  Mismatches 1181;  Indels   29;  Gaps    9;

Qy       342 GTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGANTA 401
Db        46 GAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTG 105

Qy       402 AAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAG 461
Db       106 ACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGGAA 165

Qy       462 AATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTAT 521
Db       166 GCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCAT 225

Qy       522 CTATTCGACGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTC 581
Db       226 GCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTT 285

Qy       582 AGCGGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATA 641
Db       286 TTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATGAACTT 345

Qy       642 TCGTGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCGATGAAG 701
Db       346 CCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAA 405

Qy       702 GGCCTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATA 761
Db       406 ATCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCAATTGATATT 465

Qy       762 AGGGAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAA 821
Db       466 CCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAAC 525

Qy       822 GATATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGAC 881
Db       526 AGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGAT 585

Qy       882 ATAATGCGACGTTGCTCACACAATCCGGCTCGTCATCAGGACGAAGTACGGGT------A 935
Db       586 TTCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAA 645

Qy       936 AATGTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGAT 995
Db       646 GAACTAATAGGTTTAGTTGTTCTTACCAAGTATAACAATAAGACATACAGAGTGGATGAT 705

Qy       996 GTCGACTTTGGACAAACTCCGAAATCAACATT---CAGTTGCAAGGGTAGAGATATCAGT 1052
Db       706 ATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGAAGTCAGC 765

Qy      1053 TTCGTGGAATACTATCTCACTAAATATAATATACGCATTCGCGACCACAATCAGCCGCTG 1112
Db       766 TTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTC 825

Qy      1113 CTGATTTCCAA---AAATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTA 1169
Db       826 TTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATG 885

Qy      1170 CTAATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCAG 1229
Db       886 CTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAAC 945

Qy      1230 CTTATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATCGATTG 1289
Db       946 GTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTG 1005

Qy      1290 CGCGCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGACTGG 1349
Db      1006 GGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGG 1065

Qy      1350 AACATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCAGAAC 1409
Db      1066 GGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAG 1125

Qy      1410 ATCGTGTTTCATAATGGAAAGGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTC 1469
Db      1126 AT----TCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCCAA 1181

Qy      1470 AGAGACCAAAGGATGCTTACCACTCCGAGCGATGGC-CTCGATCGTTGGGCTGTCATCGC 1528
Db      1182 AGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTA 1241

Qy      1529 GCCGCAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAG 1588
Db      1242 TACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACC 1301

Qy      1589 TGGAATGGGTCTTAGAATTCGAAGCCCCCAGGAATTCATAATTTATGATGATCGCACTGG 1648
```

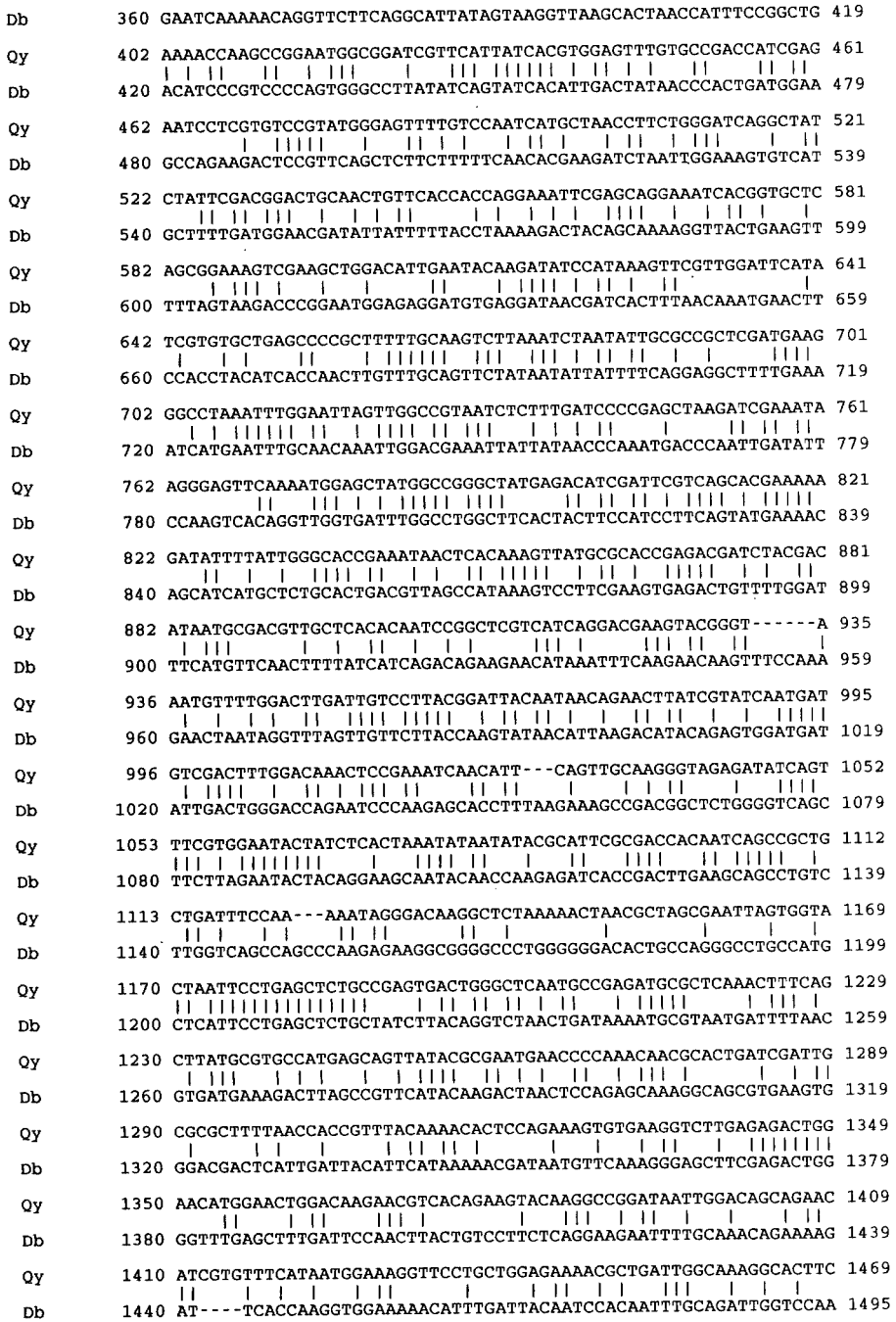
```
Db       1302 AGCCATGGGCATGCAAATGAGAAAAGC---AATAATGATTGAAGTGGATGACAGAACTGA 1358
Qy       1649 AACTTATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGCCT 1708
Db       1359 AGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCT 1418
Qy       1709 CGTACCCAATGATAACGCCGAAAGATACTCATCAATCAAAAAGAGAGGATACGTTGACAG 1768
Db       1419 GTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTG 1478
Qy       1769 GGCGGTGCCAACTCA---AGTTGTGACCCTTAAAACGACCAAGAACCGTAGCCTTATGAG 1825
Db       1479 CCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGC 1538
Qy       1826 CATTGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGAT 1885
Db       1539 CATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGT 1598
Qy       1886 CGAACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGA 1945
Db       1599 GGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGC 1658
Qy       1946 TCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTA 2005
Db       1659 TGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTT 1718
Qy       2006 CTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGAT 2065
Db       1719 CTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCT 1778
Qy       2066 GATAGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGT 2125
Db       1779 GCAAGC---GGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCAT 1835
Qy       2126 ATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAA 2185
Db       1836 CGTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCC 1895
Qy       2186 GGACATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATT 2245
Db       1896 ACAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGACTAACGGTAAT 1955
Qy       2246 AGCT---TATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAA 2302
Db       1956 TGTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAA 2015
Qy       2303 TCCTCCGCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTA 2362
Db       2016 TCCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTT 2075
Qy       2363 TCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTA 2422
Db       2076 TATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTA 2135
Qy       2423 TAGCAGCATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTT 2482
Db       2136 TGACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACAT 2195
Qy       2483 GTACTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCT 2542
Db       2196 CTATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCT 2255
Qy       2543 AGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTT 2602
Db       2256 GGCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCT 2315
Qy       2603 TTATTATCTATAA 2615
Db       2316 TTACTACCTCTAA 2328

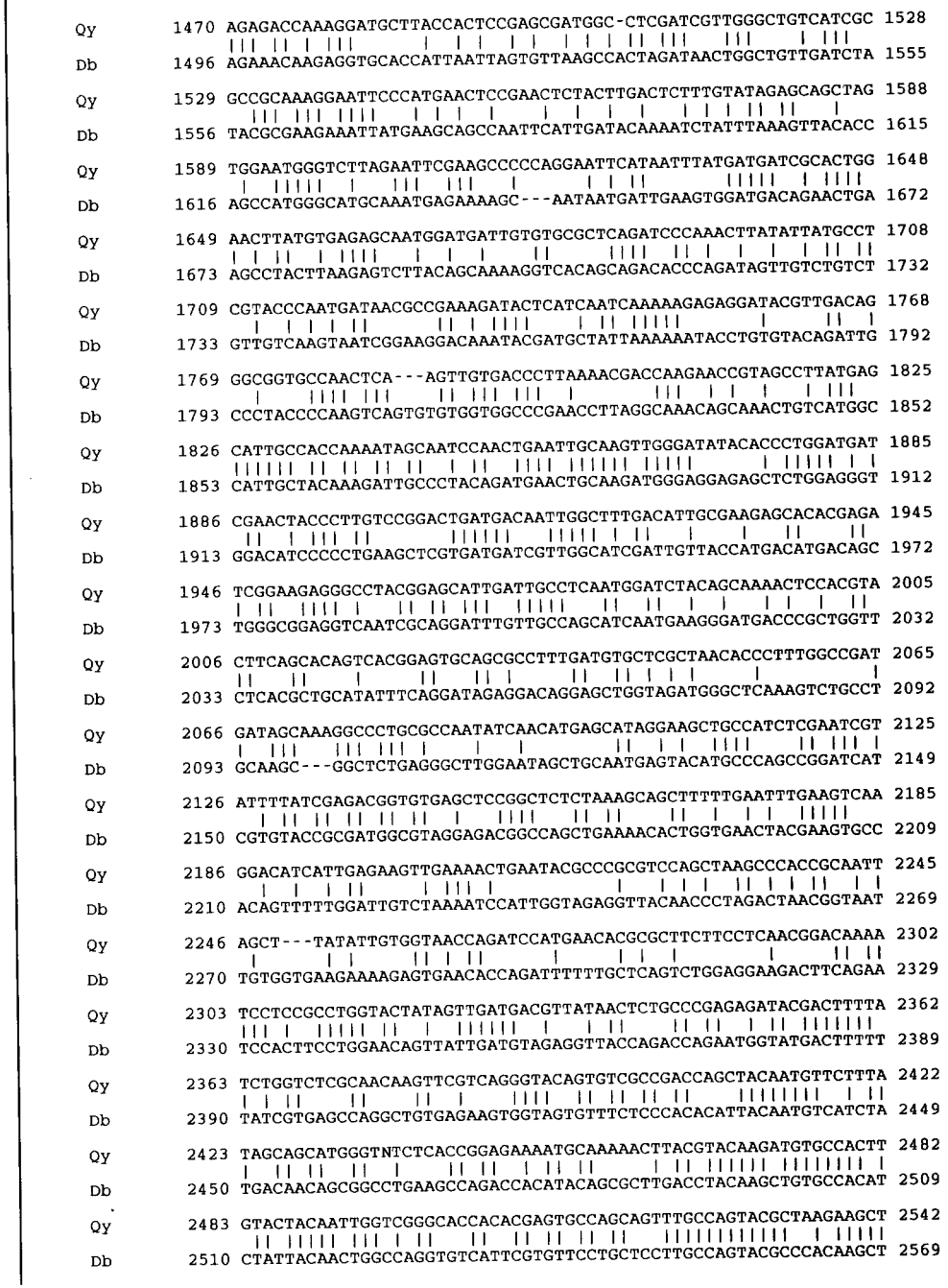
RESULT 14
```

```
AF104260
LOCUS       AF104260               3362 bp    mRNA    linear   PRI 09-JAN-2002
DEFINITION  Homo sapiens HIWI mRNA, complete cds.
ACCESSION   AF104260
VERSION     AF104260.2  GI:18098557
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3362)
  AUTHORS   Cox,D.N., Chao,A., Baker,J., Chang,L., Qiao,D. and Lin,H.
  TITLE     A novel class of evolutionarily conserved genes defined by piwi are
            essential for stem cell self-renewal
  JOURNAL   Genes Dev. 12 (23), 3715-3727 (1998)
  MEDLINE   99069219
   PUBMED   9851978
REFERENCE   2  (bases 1 to 3362)
  AUTHORS   Qiao,D., Zeeman,A.-M., Deng,W., Looijenga,L.H.J. and Lin,H.
  TITLE     Molecular characterization of hiwi, a human member of the piwi stem
            cell gene family whose overexpression is correlated to seminomas
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 3362)
  AUTHORS   Lin,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1998) Cell Biology, Duke University Medical
            Center, 412 Nanaline Duke Bldg., Research Dr., Durham, NC 27710,
            USA
REFERENCE   4  (bases 1 to 3362)
  AUTHORS   Qiao,D., Zeeman,A.-M., Deng,W., Looijenga,L.H.J. and Lin,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Cell Biology, Duke University, 412 Nanaline
            Duke Building, DUMC, Durham, NC 27710, USA
  REMARK    Nucleotide sequence updated by submitter
COMMENT     On Jan 9, 2002 this sequence version replaced gi:4038412.
FEATURES             Location/Qualifiers
     source          1. .3362
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /map="12q24.33"
                     /tissue_type="testis"
     CDS             57. .2642
                     /note="member of the PIWI family"
                     /codon_start=1
                     /product="HIWI"
                     /protein_id="AAC97371.2"
                     /db_xref="GI:18098558"
                     /translation="MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEG
                     ELFGRGRQRGTAGGTAKSQGLQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVKE
                     SKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKC
                     HAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRR
                     LLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRS
                     ETVLDFMFNFYHQTEEHKFQEQVSKELIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKK
                     ADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGL
                     TDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLL
                     SFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA
                     NSLIQNLFKVTPAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKD
                     KYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLK
                     LVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAA
                     LRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVV
                     KKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIY
                     DNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSN
                     RLYYL"
BASE COUNT     1026 a    659 c    759 g    918 t
ORIGIN

  Query Match           8.3%;  Score 253.4;  DB 9;  Length 3362;
  Best Local Similarity  47.1%;  Pred. No. 2.7e-44;
  Matches 1081;  Conservative    0;  Mismatches 1183;  Indels   29;  Gaps    9;

Qy       342 GTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGANTA 401
```





```
Qy       2543 AGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTT 2602
Db       2570 GGCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCT 2629

Qy       2603 TTATTATCTATAA 2615
Db       2630 TTACTACCTCTAA 2642
```

RESULT 15
BC028581
LOCUS BC028581 3399 bp mRNA linear PRI 25-APR-2002
DEFINITION Homo sapiens, piwi-like 1 (Drosophila), clone MGC:26748 IMAGE:4827445, mRNA, complete cds.
ACCESSION BC028581
VERSION BC028581.1 GI:20306781
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3399)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (23-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbiology.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 34 Row: b Column: 20
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.

FEATURES Location/Qualifiers
source 1. .3399
/organism="Homo sapiens"
/db_xref="LocusID:9271"
/db_xref="taxon:9606"
/clone="MGC:26748 IMAGE:4827445"
/tissue_type="Testis"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"
CDS 74. .2659
/codon_start=1
/product="piwi-like 1 (Drosophila)"
/protein_id="AAH28581.1"
/db_xref="GI:20306782"
/translation="MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEG
ELFGRGRQRGTAGGTAKSQGLQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVKE
SKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKC
HAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRR
LLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRS
ETVLDFMFNFYHQTEEHKFQEQVSKELIGLVVLTKYNNKTYRVDDIDWDQNPKSTFKK
ADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGL
TDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLL
SFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVNPLDNWLLIYTRRNYEAA
NSLIQNLFKVTPAMGMQMKKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKD
KYDAIKKYPCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLK
LVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAA
LRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVV
KKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIY
DNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSN
RLYYL"
BASE COUNT 1048 a 662 c 766 g 923 t
ORIGIN

Query Match 8.3%; Score 253.4; DB 9; Length 3399;
Best Local Similarity 47.1%; Pred. No. 2.7e-44;
Matches 1081; Conservative 0; Mismatches 1183; Indels 29; Gaps 9;

```
Qy        342 GTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGANTA 401
Db        377 GAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTG 436

Qy        402 AAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAG 461
Db        437 ACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGGAA 496

Qy        462 AATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTAT 521
Db        497 GCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCAT 556

Qy        522 CTATTCGACGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTC 581
Db        557 GCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTT 616

Qy        582 AGCGGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATA 641
Db        617 TTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATGAACTT 676

Qy        642 TCGTGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCGATGAAG 701
Db        677 CCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAA 736

Qy        702 GGCCTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATA 761
Db        737 ATCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCAATTGATATT 796

Qy        762 AGGGAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAA 821
Db        797 CCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAAC 856

Qy        822 GATATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGAC 881
Db        857 AGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGAT 916

Qy        882 ATAATGCGACGTTGCTCACACAATCCGGCTCGTCATCAGGACGAAGTACGGGT------A 935
Db        917 TTCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAA 976

Qy        936 AATGTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGAT 995
Db        977 GAACTAATAGGTTTAGTTGTTCTTACCAAGTATAACAATAAGACATACAGAGTGGATGAT 1036

Qy        996 GTCGACTTTGGACAAACTCCGAAATCAACATT---CAGTTGCAAGGGTAGAGATATCAGT 1052
Db       1037 ATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGAAGTCAGC 1096

Qy       1053 TTCGTGGAATACTATCTCACTAAATATAATATACGCATTCGCGACCACAATCAGCCGCTG 1112
Db       1097 TTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTC 1156

Qy       1113 CTGATTTCCAA---AAATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTA 1169
Db       1157 TTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATG 1216

Qy       1170 CTAATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCAG 1229
Db       1217 CTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAAC 1276

Qy       1230 CTTATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATGCGATTG 1289
```

```
Db       1277 GTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTG 1336

Qy       1290 CGCGCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGACTGG 1349

Db       1337 GGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGG 1396

Qy       1350 AACATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCAGAAC 1409

Db       1397 GGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAG 1456

Qy       1410 ATCGTGTTTCATAATGGAAAGGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTC 1469

Db       1457 AT----TCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCCAA 1512

Qy       1470 AGAGACCAAAGGATGCTTACCACTCCGAGCGATGGC-CTCGATCGTTGGGCTGTCATCGC 1528

Db       1513 AGAAACAAGAGGTGCACCATTAATTAGTGTTAATCCACTAGATAACTGGCTGTTGATCTA 1572

Qy       1529 GCCGCAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAG 1588

Db       1573 TACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACC 1632

Qy       1589 TGGAATGGGTCTTAGAATTCGAAGCCCCCAGGAATTCATAATTTATGATGATCGCACTGG 1648

Db       1633 AGCCATGGGCATGCAAATGAAAAAAGCAAT---AATGATTGAAGTGGATGACAGAACTGA 1689

Qy       1649 AACTTATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGCCT 1708

Db       1690 AGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCT 1749

Qy       1709 CGTACCCAATGATAACGCCGAAAGATACTCATCAATCAAAAAGAGAGGATACGTTGACAG 1768

Db       1750 GTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCCGTGTACAGATTG 1809

Qy       1769 GGCGGTGCCAACTCA---AGTTGTGACCCTTAAAACGACCAAGAACCGTAGCCTTATGAG 1825

Db       1810 CCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGC 1869

Qy       1826 CATTGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGAT 1885

Db       1870 CATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGT 1929

Qy       1886 CGAACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGA 1945

Db       1930 GGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGC 1989

Qy       1946 TCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTA 2005

Db       1990 TGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTT 2049

Qy       2006 CTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGAT 2065

Db       2050 CTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCT 2109

Qy       2066 GATAGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGT 2125

Db       2110 GCAAGC---GGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCAT 2166

Qy       2126 ATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAA 2185

Db       2167 CGTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCC 2226

Qy       2186 GGACATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATT 2245

Db       2227 ACAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGACTAACGGTAAT 2286

Qy       2246 AGCT---TATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAA 2302

Db       2287 TGTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAA 2346

Qy       2303 TCCTCCGCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTA 2362

Db       2347 TCCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTT 2406

Qy       2363 TCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTA 2422

Db       2407 TATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTA 2466

Qy       2423 TAGCAGCATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTT 2482

Db       2467 TGACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACAT 2526

Qy       2483 GTACTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCT 2542

Db       2527 CTATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCT 2586

Qy       2543 AGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTT 2602

Db       2587 GGCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCT 2646

Qy       2603 TTATTATCTATAA 2615

Db       2647 TTACTACCTCTAA 2659
```

Search completed: July 22, 2003, 06:18:31
Job time : 7730.12 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2003, 11:38:12 ; Search time 607.788 Seconds
(without alignments)
11289.862 Million cell updates/sec

Title: US-09-873-737A-1
Perfect score: 3047
Sequence: 1 ctgagtccaaagcgtcgttt.....aaaaaaaaaaaaaaaaaaa 3047

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3044	99.9	3047	21	AAA07586	Drosophila piwi ge
2	2980.2	97.8	3524	23	ABL10535	Drosophila melanog
3	1734.4	56.9	7667	23	ABL10534	Drosophila melanog
4	479.4	15.7	2838	23	ABL10571	Drosophila melanog
5	389	12.8	6886	23	ABL08020	Drosophila melanog
6	311.4	10.2	4064	21	AAA07587	Mouse piwi gene, d
7	253	8.3	3472	21	AAA07588	Human piwi gene, d
8	221.6	7.3	5671	23	ABL10570	Drosophila melanog
9	167.2	5.5	3423	23	ABL08096	Drosophila melanog
10	164	5.4	3649	22	AAS13630	cDNA sequence enco
11	125.8	4.1	2272	22	AAH15960	Human cDNA sequenc
12	101.8	3.3	1733	22	AAH14032	Human cDNA sequenc
13	81.6	2.7	646	23	AAS77101	DNA encoding novel
c 14	77.2	2.5	449	23	ABV56647	Human prostate exp
15	76.6	2.5	2750	14	AAQ38808	G6PD coding sequen
16	76.2	2.5	570	23	ABV59114	Human prostate exp
c 17	75.8	2.5	6070	24	ABL92199	Chemically treated
c 18	75.8	2.5	6070	24	ABL49310	Human polynucleoti
c 19	75.8	2.5	6070	24	ABL32241	Human immune syste
20	75.4	2.5	406	22	AAI83618	Human polynucleoti
c 21	75.2	2.5	6145	24	ABL32972	Human immune syste
c 22	74.6	2.4	300	22	AAH70049	Human cervical can
23	74.4	2.4	413	22	AAI85186	Human polynucleoti
c 24	73.8	2.4	612	22	AAH71471	Human cervical can
c 25	73.6	2.4	8446	24	ABL33671	Human immune syste
c 26	73.4	2.4	893	23	ABV04793	Human prostate exp
27	73.4	2.4	1046	21	AAC99079	Human pancreatic c
28	73	2.4	309	23	ABV44994	Human prostate exp
29	73	2.4	2103	22	AAD08708	Lycopersicon penne
30	73	2.4	2457	22	AAS33081	DNA encoding human
c 31	72.8	2.4	509	22	AAS60503	Human cancer agent
32	72.8	2.4	981	22	AAS11352	Corn DNA encoding
33	72.8	2.4	1026	23	AAS70144	DNA encoding novel
c 34	72.6	2.4	6327	22	AAS45444	Chemically pretrea
c 35	72.6	2.4	6327	24	ABK28293	DNA transcription
36	72.4	2.4	673	22	AAH34055	Human colon cancer
c 37	72.4	2.4	695	22	AAL15586	Human breast cance
c 38	72.4	2.4	6045	24	ABL70623	Chemically treated
c 39	72.4	2.4	6045	24	ABK31540	Signal transductio
40	72.2	2.4	250	22	AAS29153	cDNA encoding for
c 41	72.2	2.4	327	23	ABV06120	Human prostate exp
42	72.2	2.4	1194	22	AAH34600	Human colon cancer
43	72.2	2.4	1336	22	AAS29081	cDNA encoding for
44	72.2	2.4	3366	22	AAF24684	Nucleotide sequenc
45	72.2	2.4	3366	22	AAF24706	Nucleotide sequenc

ALIGNMENTS

RESULT 1
AAA07586
ID AAA07586 standard; DNA; 3047 BP.
XX
AC AAA07586;
XX
DT 29-AUG-2000 (first entry)
XX
DE Drosophila piwi gene.
XX
KW Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW anaemia; immunodeficiency; male infertility; Drosophila; ds.
XX
OS Drosophila sp.
XX
FH Key Location/Qualifiers
FT CDS 84..2615
FT /*tag= a
FT /product= piwi
FT /transl_except= (pos:120..122; aa:Xaa)
FT /transl_except= (pos:399..401; aa:Xaa)
FT /transl_except= (pos:2436..2438; aa:Xaa)
FT /note= "Xaa= Leu or Ile"
XX
PN WO200032039-A1.
XX
PD 08-JUN-2000.
XX
PF 03-DEC-1999; 99WO-US28764.
XX
PR 04-DEC-1998; 98US-0110901.
XX

PA (UYDU-) UNIV DUKE.
XX
PI Lin H;
XX
DR WPI; 2000-412085/35.
DR P-PSDB; AAY90233.
XX
PT Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT therapy of diseases such as cancer and in various research and
PT diagnostic applications -
XX
PS Claim 19; Page 171-175; 201pp; English.
XX
CC This sequence encodes the Drosophila piwi family protein, designated
CC piwi. The piwi family nucleic acids and polypeptides are used in gene
CC therapy of diseases such as cancer and also in various research and
CC diagnostic applications. The sequences can also be used to treat
CC tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ Sequence 3047 BP; 936 A; 664 C; 672 G; 772 T; 3 other;

Query Match 99.9%; Score 3044; DB 21; Length 3047;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3047; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CTGAGTCCAAAGCGTCGTTTTCAAAGTACTCTTTCAGTTTCCATTGTGAAGTTTTAAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTGAGTCCAAAGCGTCGTTTTCAAAGTACTCTTTCAGTTTCCATTGTGAAGTTTTAAGTG 60

Qy         61 ATCGCGAGTGCCAAAAAGTAACAATGGCTGATGATCAGGGACGTGGACGCAGGCGTCCAN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATCGCGAGTGCCAAAAAGTAACAATGGCTGATGATCAGGGACGTGGACGCAGGCGTCCAN 120

Qy        121 TTAACGAAGATGATTCCTCTACTTCCCGAGGTAGTGGTGATGGGCCGCGGGTGAAAGTAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAACGAAGATGATTCCTCTACTTCCCGAGGTAGTGGTGATGGGCCGCGGGTGAAAGTAT 180

Qy        181 TCAGAGGATCTTCATCAGGTGACCCGAGAGCGGATCCTCGTATAGAGGCTTCAAGAGAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCAGAGGATCTTCATCAGGTGACCCGAGAGCGGATCCTCGTATAGAGGCTTCAAGAGAGA 240

Qy        241 GAAGAGCTCTCGAGGAAGCTCCCAGGCGTGAAGGTGGCCCGCCAGAGCGAAAGCCGTGGG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAAGAGCTCTCGAGGAAGCTCCCAGGCGTGAAGGTGGCCCGCCAGAGCGAAAGCCGTGGG 300

Qy        301 GTGACCAATATGATTACCTGAATACCCGTCCGGTTGAGCTGGTATCCAAGAAGGGAACCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTGACCAATATGATTACCTGAATACCCGTCCGGTTGAGCTGGTATCCAAGAAGGGAACCG 360

Qy        361 ATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTTCGANTAAAAACCAAGCCGGAATGGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTTCGANTAAAAACCAAGCCGGAATGGC 420

Qy        421 GGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAGAATCCTCGTGTCCGTATGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAGAATCCTCGTGTCCGTATGG 480

Qy        481 GAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTATCTATTCGACGGACTGCAAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTATCTATTCGACGGACTGCAAC 540

Qy        541 TGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTCAGCGGAAAGTCGAAGCTGG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTCAGCGGAAAGTCGAAGCTGG 600

Qy        601 ACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATATCGTGTGCTGAGCCCCGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATATCGTGTGCTGAGCCCCGCT 660

Qy        661 TTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCGATGAAGGGCCTAAATTTGGAATTAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCGATGAAGGGCCTAAATTTGGAATTAG 720

Qy        721 TTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATAAGGGAGTTCAAAATGGAGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATAAGGGAGTTCAAAATGGAGC 780

Qy        781 TATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAAGATATTTTATTGGGCACCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAAGATATTTTATTGGGCACCG 840

Qy        841 AAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGACATAATGCGACGTTGCTCAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGACATAATGCGACGTTGCTCAC 900

Qy        901 ACAATCCGGCTCGTCATCAGGACGAAGTACGGGTAAATGTTTTGGACTTGATTGTCCTTA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACAATCCGGCTCGTCATCAGGACGAAGTACGGGTAAATGTTTTGGACTTGATTGTCCTTA 960

Qy        961 CGGATTACAATAACAGAACTTATCGTATCAATGATGTCGACTTTGGACAAACTCCGAAAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CGGATTACAATAACAGAACTTATCGTATCAATGATGTCGACTTTGGACAAACTCCGAAAT 1020

Qy       1021 CAACATTCAGTTGCAAGGGTAGAGATATCAGTTTCGTGGAATACTATCTCACTAAATATA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAACATTCAGTTGCAAGGGTAGAGATATCAGTTTCGTGGAATACTATCTCACTAAATATA 1080

Qy       1081 ATATACGCATTCGCGACCACAATCAGCCGCTGCTGATTTCCAAAAATAGGGACAAGGCTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ATATACGCATTCGCGACCACAATCAGCCGCTGCTGATTTCCAAAAATAGGGACAAGGCTC 1140

Qy       1141 TAAAAACTAACGCTAGCGAATTAGTGGTACTAATTCCTGAGCTCTGCCGAGTGACTGGGC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TAAAAACTAACGCTAGCGAATTAGTGGTACTAATTCCTGAGCTCTGCCGAGTGACTGGGC 1200

Qy       1201 TCAATGCCGAGATGCGCTCAAACTTTCAGCTTATGCGTGCCATGAGCAGTTATACGCGAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCAATGCCGAGATGCGCTCAAACTTTCAGCTTATGCGTGCCATGAGCAGTTATACGCGAA 1260

Qy       1261 TGAACCCCAAACAACGCACTGATCGATTGCGCGCTTTTAACCACCGTTTACAAAACACTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TGAACCCCAAACAACGCACTGATCGATTGCGCGCTTTTAACCACCGTTTACAAAACACTC 1320

Qy       1321 CAGAAAGTGTGAAGGTCTTGAGAGACTGGAACATGGAACTGGACAAGAACGTCACAGAAG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CAGAAAGTGTGAAGGTCTTGAGAGACTGGAACATGGAACTGGACAAGAACGTCACAGAAG 1380

Qy       1381 TACAAGGCCGGATAATTGGACAGCAGAACATCGTGTTTCATAATGGAAAGGTTCCTGCTG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TACAAGGCCGGATAATTGGACAGCAGAACATCGTGTTTCATAATGGAAAGGTTCCTGCTG 1440

Qy       1441 GAGAAAACGCTGATTGGCAAAGGCACTTCAGAGACCAAAGGATGCTTACCACTCCGAGCG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGAAAACGCTGATTGGCAAAGGCACTTCAGAGACCAAAGGATGCTTACCACTCCGAGCG 1500

Qy       1501 ATGGCCTCGATCGTTGGGCTGTCATCGCGCCGCAAAGGAATTCCCATGAACTCCGAACTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 ATGGCCTCGATCGTTGGGCTGTCATCGCGCCGCAAAGGAATTCCCATGAACTCCGAACTC 1560

Qy       1561 TACTTGACTCTTTGTATAGAGCAGCTAGTGGAATGGGTCTTAGAATTCGAAGCCCCCAGG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TACTTGACTCTTTGTATAGAGCAGCTAGTGGAATGGGTCTTAGAATTCGAAGCCCCCAGG 1620

Qy       1621 AATTCATAATTTATGATGATCGCACTGGAACTTATGTGAGAGCAATGGATGATTGTGTGC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AATTCATAATTTATGATGATCGCACTGGAACTTATGTGAGAGCAATGGATGATTGTGTGC 1680

Qy       1681 GCTCAGATCCCAAACTTATATTATGCCTCGTACCCAATGATAACGCCGAAAGATACTCAT 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GCTCAGATCCCAAACTTATATTATGCCTCGTACCCAATGATAACGCCGAAAGATACTCAT 1740

Qy       1741 CAATCAAAAAGAGAGGATACGTTGACAGGGCGGTGCCAACTCAAGTTGTGACCCTTAAAA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CAATCAAAAAGAGAGGATACGTTGACAGGGCGGTGCCAACTCAAGTTGTGACCCTTAAAA 1800

Qy       1801 CGACCAAGAACCGTAGCCTTATGAGCATTGCCACCAAAATAGCAATCCAACTGAATTGCA 1860
```

```
Db       1801 CGACCAAGAACCGTAGCCTTATGAGCATTGCCACCAAAATAGCAATCCAACTGAATTGCA 1860

Qy       1861 AGTTGGGATATACACCCTGGATGATCGAACTACCCTTGTCCGGACTGATGACAATTGGCT 1920

Db       1861 AGTTGGGATATACACCCTGGATGATCGAACTACCCTTGTCCGGACTGATGACAATTGGCT 1920

Qy       1921 TTGACATTGCGAAGAGCACACGAGATCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAA 1980

Db       1921 TTGACATTGCGAAGAGCACACGAGATCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAA 1980

Qy       1981 TGGATCTACAGCAAAACTCCACGTACTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATG 2040

Db       1981 TGGATCTACAGCAAAACTCCACGTACTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATG 2040

Qy       2041 TGCTCGCTAACACCCTTTGGCCGATGATAGCAAAGGCCCTGCGCCAATATCAACATGAGC 2100

Db       2041 TGCTCGCTAACACCCTTTGGCCGATGATAGCAAAGGCCCTGCGCCAATATCAACATGAGC 2100

Qy       2101 ATAGGAAGCTGCCATCTCGAATCGTATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAA 2160

Db       2101 ATAGGAAGCTGCCATCTCGAATCGTATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAA 2160

Qy       2161 AGCAGCTTTTTGAATTTGAAGTCAAGGACATCATTGAGAAGTTGAAAACTGAATACGCCC 2220

Db       2161 AGCAGCTTTTTGAATTTGAAGTCAAGGACATCATTGAGAAGTTGAAAACTGAATACGCCC 2220

Qy       2221 GCGTCCAGCTAAGCCCACCGCAATTAGCTTATATTGTGGTAACCAGATCCATGAACACGC 2280

Db       2221 GCGTCCAGCTAAGCCCACCGCAATTAGCTTATATTGTGGTAACCAGATCCATGAACACGC 2280

Qy       2281 GCTTCTTCCTCAACGGACAAAATCCTCCGCCTGGTACTATAGTTGATGACGTTATAACTC 2340

Db       2281 GCTTCTTCCTCAACGGACAAAATCCTCCGCCTGGTACTATAGTTGATGACGTTATAACTC 2340

Qy       2341 TGCCCGAGAGATACGACTTTTATCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGC 2400

Db       2341 TGCCCGAGAGATACGACTTTTATCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGC 2400

Qy       2401 CGACCAGCTACAATGTTCTTTATAGCAGCATGGGTNTCTCACCGGAGAAAATGCAAAAAC 2460

Db       2401 CGACCAGCTACAATGTTCTTTATAGCAGCATGGGTNTCTCACCGGAGAAAATGCAAAAAC 2460

Qy       2461 TTACGTACAAGATGTGCCACTTGTACTACAATTGGTCGGGCACCACACGAGTGCCAGCAG 2520

Db       2461 TTACGTACAAGATGTGCCACTTGTACTACAATTGGTCGGGCACCACACGAGTGCCAGCAG 2520

Qy       2521 TTTGCCAGTACGCTAAGAAGCTAGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGC 2580

Db       2521 TTTGCCAGTACGCTAAGAAGCTAGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGC 2580

Qy       2581 AAAACGCGCTCGAAAAGAAGTTTTATTATCTATAATTGGATATAATTTAGAATGGAGTAT 2640

Db       2581 AAAACGCGCTCGAAAAGAAGTTTTATTATCTATAATTGGATATAATTTAGAATGGAGTAT 2640

Qy       2641 TAATCCTTACTAAGAGGCCATATATGAAACTAGCCCAGACATTTATACTTTTTCAATACT 2700

Db       2641 TAATCCTTACTAAGAGGCCATATATGAAACTAGCCCAGACATTTATACTTTTTCAATACT 2700

Qy       2701 TCCTTACTTTTGCTAAGCACTTCAGCATTTATGACTAAATATTTTGTATTTGAAATGCAT 2760

Db       2701 TCCTTACTTTTGCTAAGCACTTCAGCATTTATGACTAAATATTTTGTATTTGAAATGCAT 2760

Qy       2761 TACTGCTCTTTTTTCAAACAAAAGCAAAATTGAGGATTAAGATTCTGGTATTTAAGCATA 2820

Db       2761 TACTGCTCTTTTTTCAAACAAAAGCAAAATTGAGGATTAAGATTCTGGTATTTAAGCATA 2820

Qy       2821 AGACCAGAGGAAATTCCCAAACAAACATTTAAAGTTATCTATCAAGACATGTTCATTAAT 2880

Db       2821 AGACCAGAGGAAATTCCCAAACAAACATTTAAAGTTATCTATCAAGACATGTTCATTAAT 2880

Qy       2881 TTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTTTATGTAAAGAAAAATTACA 2940

Db       2881 TTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTTTATGTAAAGAAAAATTACA 2940

Qy       2941 TACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGCTTATATTCGTGAAAAAAAA 3000

Db       2941 TACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGCTTATATTCGTGAAAAAAAA 3000

Qy       3001 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3047

Db       3001 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3047
```

RESULT 2
ABL10535
ID ABL10535 standard; cDNA; 3524 BP.
XX
AC ABL10535;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 26087.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB66432.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 26087; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3524 BP; 1053 A; 741 C; 761 G; 969 T; 0 other;

Query Match 97.8%; Score 2980.2; DB 23; Length 3524;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 3003; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

```
Qy          1 CTGAGTCCAAAGCGTCGTTTTCAAAGTACTCTTTCAGTTTCCATTGTGAAGTTTTAAGTG 60

Db          1 CTGAGTCCAAAGCGTCGTTTTCAAAGTACTCTTTCAGTTTCCATTGTGAAGTTTTAAGTG 60

Qy         61 ATCGCGAGTGCCAAAAAGTAACAATGGCTGATGATCAGGGACGTGGACGCAGGCGTCCAN 120

Db         61 ATCGCGAGTGCCAAAAAGTAACAATGGCTGATGATCAGGGACGTGGACGCAGGCGTCCAC 120
```

```
Qy      121 TTAACGAAGATGATTCCTCTACTTCCCGAGGTAGTGGTGATGGGCCGCGGGTGAAAGTAT 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 TTAACGAAGATGATTCCTCTACTTCCCGAGGTAGTGGTGATGGGCCGCGGGTGAAAGTAT 180

Qy      181 TCAGAGGATCTTCATCAGGTGACCCGAGAGCGGATCCTCGTATAGAGGCTTCAAGAGAGA 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 TCAGAGGATCTTCATCAGGTGACCCGAGAGCGGATCCTCGTATAGAGGCTTCAAGAGAGA 240

Qy      241 GAAGAGCTCTCGAGGAAGCTCCCAGGCGTGAAGGTGGCCCGCCAGAGCGAAAGCCGTGGG 300
            ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db      241 GAAGAGCTCTCGAGGAAGCTCCCAGGCGTGAAGGTGGCCCGACAGAGCGAAAGCCGTGGG 300

Qy      301 GTGACCAATATGATTACCTGAATACCCGTCCGGTTGAGCTGGTATCCAAGAAGGGAACCG 360
            |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db      301 GTGACCAATATGATTACCTGAATACCCGTCCGGCTGAGCTGGTATCCAAGAAGGGAACCG 360

Qy      361 ATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGANTAAAAACCAAGCCGGAATGGC 420
            |||||||||||||||||||||||||||||| ||| ||||||||||||||||||||||||||
Db      361 ATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTCCGATTAAAAACCAAGCCGGAATGGC 420

Qy      421 GGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAGAATCCTCGTGTCCGTATGG 480
            |||||||||||||||||||||||||| ||||| ||||||||||||||||||||||||||||
Db      421 GGATCGTTCATTATCACGTGGAGTTTGAGCCGAGCATCGAGAATCCTCGTGTCCGTATGG 480

Qy      481 GAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTATCTATTCGACGGACTGCAAC 540
            |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      481 GAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTATCTATTCGATGGACTGCAAC 540

Qy      541 TGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTCAGCGGAAAGTCGAAGCTGG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 TGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTCAGCGGAAAGTCGAAGCTGG 600

Qy      601 ACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATATCGTGTGCTGAGCCCCGCT 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 ACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATATCGTGTGCTGAGCCCCGCT 660

Qy      661 TTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCGATGAAGGGCCTAAATTTGGAATTAG 720
            |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db      661 TTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCCATGAAGGGCCTAAATTTGGAATTAG 720

Qy      721 TTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATAAGGGAGTTCAAAATGGAGC 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 TTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATAAGGGAGTTCAAAATGGAGC 780

Qy      781 TATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAAGATATTTTATTGGGCACCG 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 TATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAAGATATTTTATTGGGCACCG 840

Qy      841 AAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGACATAATGCGACGTTGCTCAC 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 AAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGACATAATGCGACGTTGCTCAC 900

Qy      901 ACAATCCGGCTCGTCATCAGGACGAAGTACGGGTAAATGTTTTGGACTTGATTGTCCTTA 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 ACAATCCGGCTCGTCATCAGGACGAAGTACGGGTAAATGTTTTGGACTTGATTGTCCTTA 960

Qy      961 CGGATTACAATAACAGAACTTATCGTATCAATGATGTCGACTTTGGACAAACTCCGAAAT 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 CGGATTACAATAACAGAACTTATCGTATCAATGATGTCGACTTTGGACAAACTCCGAAAT 1020

Qy     1021 CAACATTCAGTTGCAAGGGTAGAGATATCAGTTTCGTGGAATACTATCTCACTAAATATA 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 CAACATTCAGTTGCAAGGGTAGAGATATCAGTTTCGTGGAATACTATCTCACTAAATATA 1080

Qy     1081 ATATACGCATTCGCGACCACAATCAGCCGCTGCTGATTTCCAAAAATAGGGACAAGGCTC 1140
            |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     1081 ATATACGCATTCGCGACCACAATCAGCCGCTGCTGATCTCCAAAAATAGGGACAAGGCTC 1140

Qy     1141 TAAAAACTAACGCTAGCGAATTAGTGGTACTAATTCCTGAGCTCTGCCGAGTGACTGGGC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141 TAAAAACTAACGCTAGCGAATTAGTGGTACTAATTCCTGAGCTCTGCCGAGTGACTGGGC 1200

Qy     1201 TCAATGCCGAGATGCGCTCAAACTTTCAGCTTATGCGTGCCATGAGCAGTTATACGCGAA 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201 TCAATGCCGAGATGCGCTCAAACTTTCAGCTTATGCGTGCCATGAGCAGTTATACGCGAA 1260

Qy     1261 TGAACCCCAAACAACGCACTGATCGATTGCGCGCTTTTAACCACCGTTTACAAAACACTC 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261 TGAACCCCAAACAACGCACTGATCGATTGCGCGCTTTTAACCACCGTTTACAAAACACTC 1320

Qy     1321 CAGAAAGTGTGAAGGTCTTGAGAGACTGGAACATGGAACTGGACAAGAACGTCACAGAAG 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321 CAGAAAGTGTGAAGGTCTTGAGAGACTGGAACATGGAACTGGACAAGAACGTCACAGAAG 1380

Qy     1381 TACAAGGCCGGATAATTGGACAGCAGAACATCGTGTTTCATAATGGAAAGGTTCCTGCTG 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1381 TACAAGGCCGGATAATTGGACAGCAGAACATCGTGTTTCATAATGGAAAGGTTCCTGCTG 1440

Qy     1441 GAGAAAACGCTGATTGGCAAAGGCACTTCAGAGACCAAAGGATGCTTACCACTCCGAGCG 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1441 GAGAAAACGCTGATTGGCAAAGGCACTTCAGAGACCAAAGGATGCTTACCACTCCGAGCG 1500

Qy     1501 ATGGCCTCGATCGTTGGGCTGTCATCGCGCCGCAAAGGAATTCCCATGAACTCCGAACTC 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1501 ATGGCCTCGATCGTTGGGCTGTCATCGCGCCGCAAAGGAATTCCCATGAACTCCGAACTC 1560

Qy     1561 TACTTGACTCTTTGTATAGAGCAGCTAGTGGAATGGGTCTTAGAATTCGAAGCCCCCAGG 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1561 TACTTGACTCTTTGTATAGAGCAGCTAGTGGAATGGGTCTTAGAATTCGAAGCCCCCAGG 1620

Qy     1621 AATTCATAATTTATGATGATCGCACTGGAACTTATGTGAGAGCAATGGATGATTGTGTGC 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1621 AATTCATAATTTATGATGATCGCACTGGAACTTATGTGAGAGCAATGGATGATTGTGTGC 1680

Qy     1681 GCTCAGATCCCAAACTTATATTATGCCTCGTACCCAATGATAACGCCGAAAGATACTCAT 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1681 GCTCAGATCCCAAACTTATATTATGCCTCGTACCCAATGATAACGCCGAAAGATACTCAT 1740

Qy     1741 CAATCAAAAAGAGAGGATACGTTGACAGGGCGGTGCCAACTCAAGTTGTGACCCTTAAAA 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1741 CAATCAAAAAGAGAGGATACGTTGACAGGGCGGTGCCAACTCAAGTTGTGACCCTTAAAA 1800

Qy     1801 CGACCAAGAACCGTAGCCTTATGAGCATTGCCACCAAAATAGCAATCCAACTGAATTGCA 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1801 CGACCAAGAACCGTAGCCTTATGAGCATTGCCACCAAAATAGCAATCCAACTGAATTGCA 1860

Qy     1861 AGTTGGGATATACACCCTGGATGATCGAACTACCCTTGTCCGGACTGATGACAATTGGCT 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1861 AGTTGGGATATACACCCTGGATGATCGAACTACCCTTGTCCGGACTGATGACAATTGGCT 1920

Qy     1921 TTGACATTGCGAAGAGCACACGAGATCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAA 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1921 TTGACATTGCGAAGAGCACACGAGATCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAA 1980

Qy     1981 TGGATCTACAGCAAAACTCCACGTACTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATG 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1981 TGGATCTACAGCAAAACTCCACGTACTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATG 2040

Qy     2041 TGCTCGCTAACACCCTTTGGCCGATGATAGCAAAGGCCCTGCGCCAATATCAACATGAGC 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2041 TGCTCGCTAACACCCTTTGGCCGATGATAGCAAAGGCCCTGCGCCAATATCAACATGAGC 2100

Qy     2101 ATAGGAAGCTGCCATCTCGAATCGTATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAA 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2101 ATAGGAAGCTGCCATCTCGAATCGTATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAA 2160

Qy     2161 AGCAGCTTTTTGAATTTGAAGTCAAGGACATCATTGAGAAGTTGAAAACTGAATACGCCC 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2161 AGCAGCTTTTTGAATTTGAAGTCAAGGACATCATTGAGAAGTTGAAAACTGAATACGCCC 2220

Qy     2221 GCGTCCAGCTAAGCCCACCGCAATTAGCTTATATTGTGGTAACCAGATCCATGAACACGC 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2221 GCGTCCAGCTAAGCCCACCGCAATTAGCTTATATTGTGGTAACCAGATCCATGAACACGC 2280

Qy     2281 GCTTCTTCCTCAACGGACAAAATCCTCCGCCTGGTACTATAGTTGATGACGTTATAACTC 2340
```

```
Db       2281 GCTTCTTCCTCAACGGACAAAATCCTCCGCCTGGTACTATAGTTGATGACGTTATAACTC 2340

Qy       2341 TGCCCGAGAGATACGACTTTTATCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGC 2400

Db       2341 TGCCCGAGAGATACGACTTTTATCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGC 2400

Qy       2401 CGACCAGCTACAATGTTCTTTATAGCAGCATGGGTNTCTCACCGGAGAAAATGCAAAAAC 2460

Db       2401 CGACCAGCTACAATGTTCTTTATAGCAGCATGGGTCTCTCACCGGAGAAAATGCAAAAAC 2460

Qy       2461 TTACGTACAAGATGTGCCACTTGTACTACAATTGGTCGGGCACCACACGAGTGCCAGCAG 2520

Db       2461 TTACGTACAAGATGTGCCACTTGTACTACAATTGGTCGGGCACCACACGAGTGCCAGCAG 2520

Qy       2521 TTTGCCAGTACGCTAAGAAGCTAGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGC 2580

Db       2521 TTTGCCAGTACGCTAAGAAGCTAGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGC 2580

Qy       2581 AAAACGCGCTCGAAAAGAAGTTTTATTATCTATAATTGGATATAATTTAGAATGGAGTAT 2640

Db       2581 AAAACGCGCTCGAAAAGAAGTTTTATTATCTATAATTGGATATAATTTAGAATGGAGTAT 2640

Qy       2641 TAATCCTTACTAAGAGGCCATATATGAAACTAGCCCAGACATTTATACTTTTTCAATACT 2700

Db       2641 TAATCCTTACTAAGAGGCCATATATGAAACTAGCCCAGACATTTATACTTTTTCAATACT 2700

Qy       2701 TCCTTACTTTTGCTAAGCACTTCAGCATTTATGACTAAATATTTTGTATTTGAAATGCAT 2760

Db       2701 TTCTTACTTTTGCTAAGCACTTCAGCATTTATGACTAAATATTTTGTATTTGAAATGCAT 2760

Qy       2761 TACTGCTCTTTTTTCAAACAAAAGCAAAATTGAGGATTAAGATTCTGGTATTTAAGCATA 2820

Db       2761 TACTGCTCTTTTTTCAAACAAAAGCAAAATTGAGGATTAAGATTCTGGTATTTAAGCATA 2820

Qy       2821 AGACCAGAGGAAATTCCCAAACAAACATTTAAAGTTATCTATCAAGACATGTTCATTAAT 2880

Db       2821 AGACCAGAGGAAATTCCCAAACAAACATTTAAAGTTATCTATCAAGACATGTTCATTAAT 2880

Qy       2881 TTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTTTATGTAAAGAAAAATTACA 2940

Db       2881 TTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTTTATGTAAAGAAAAATTACA 2940

Qy       2941 TACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGCTTATATTCGTGAAAAAAAA 3000

Db       2941 TACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGCTTATATTCGTGAAACATAT 3000

Qy       3001 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3044

Db       3001 TTTATTAATTATACAGGTATTTAACTTTCACAAGGATAATACAA 3044
```

RESULT 3
ABL10534
ID ABL10534 standard; cDNA; 7667 BP.
XX
AC ABL10534;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 26084.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB66431.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 26084; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 7667 BP; 2236 A; 1513 C; 1554 G; 2364 T; 0 other;

Query Match 56.9%; Score 1734.4; DB 23; Length 7667;
Best Local Similarity 72.7%; Pred. No. 2e-304;
Matches 2845; Conservative 0; Mismatches 48; Indels 1023; Gaps 6;

```
Qy        152 TAGTGGTGATGGGCCGCGGGTGAAAGTATTCAGAGGATCTTCATCAGGTGACCCGAGAGC 211

Db       2272 TAATGCTTTCATGCAGCGGGTGAAAGTATTCAGAGGATCTTCATCAGGTGACCCGAGAGC 2331

Qy        212 GGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGAAGCTCCCAGGCGTGA 271

Db       2332 GGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGAAGCTCCCAGGCGTGA 2391

Qy        272 AGGTGGCCCGCCAGAGCGAAAGCCGTGGGGTGACCAATATGATTACCTGAATACCCGTCC 331

Db       2392 AGGTGGCCCGACAGAGCGAAAGCCGTGGGGTGACCAATATGATTACCTGAATACCCGTCC 2451

Qy        332 GGTTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTT 391

Db       2452 GGCTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTT 2511

Qy        392 TTTTCGANTAAAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCC 451

Db       2512 TTTCCGATTAAAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGAGCC 2571

Qy        452 GACCATCGAGAATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGG 511

Db       2572 GAGCATCGAGAATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGG 2631

Qy        512 ATCAGGCTATCTATTCGACGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAAT 571

Db       2632 ATCAGGCTATCTATTCGATGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAAT 2691

Qy        572 CACGGTGCTCAGCGGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGT 631

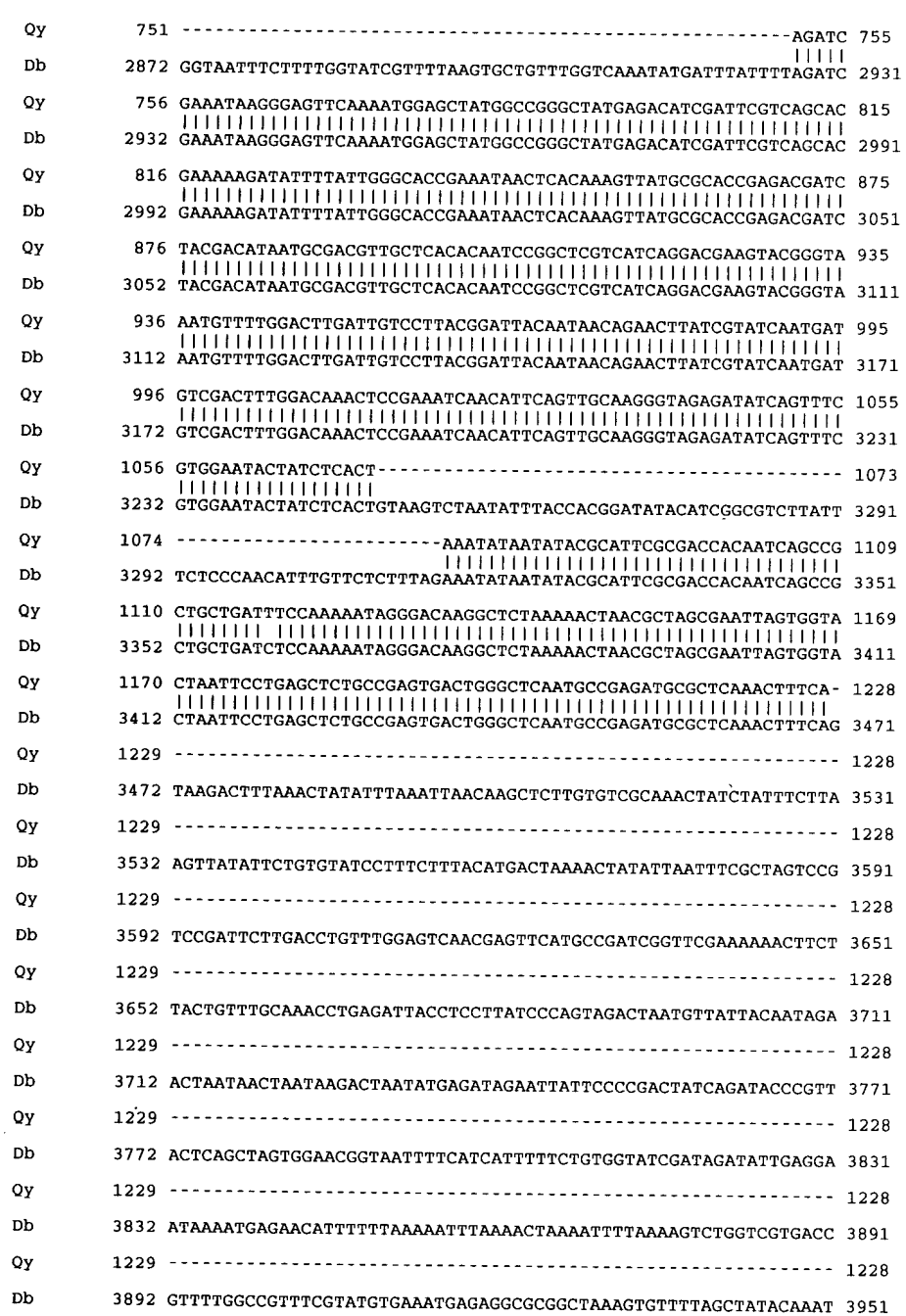
Db       2692 CACGGTGCTCAGCGGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGT 2751

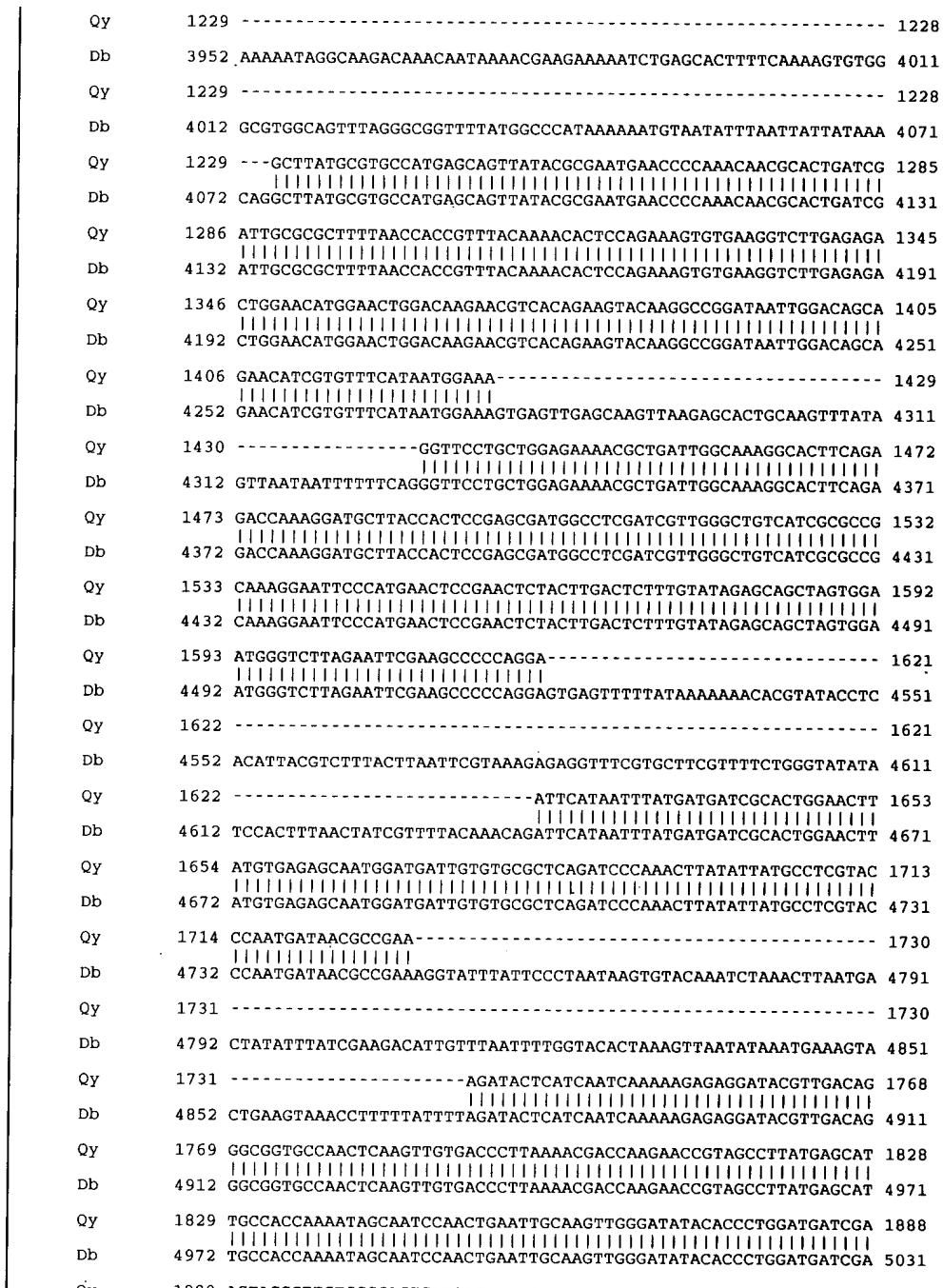
Qy        632 TGGATTCATATCGTGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCG 691

Db       2752 TGGATTCATATCGTGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCG 2811

Qy        692 CTCGATGAAGGGCCTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTA- 750

Db       2812 CTCCATGAAGGGCCTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTAA 2871
```





```
Db       5032 ACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGATCG 5091

Qy       1949 GAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTT 2008

Db       5092 GAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTT 5151

Qy       2009 CAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATGAT 2068

Db       5152 CAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATGAT 5211

Qy       2069 AGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATT 2128

Db       5212 AGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATT 5271

Qy       2129 TTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAAGGA 2188

Db       5272 TTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAAGGA 5331

Qy       2189 CATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATTAGC 2248

Db       5332 CATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATTAGC 5391

Qy       2249 TTATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAATCCTCC 2308

Db       5392 TTATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAATCCTCC 5451

Qy       2309 GCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTATCTGGT 2368

Db       5452 GCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTATCTGGT 5511

Qy       2369 CTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAGCAG 2428

Db       5512 CTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAGCAG 5571

Qy       2429 CATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTACTA 2488

Db       5572 CATGGGTCTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTACTA 5631

Qy       2489 CAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGCTAC 2548

Db       5632 CAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGCTAC 5691

Qy       2549 CCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTATTA 2608

Db       5692 CCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTATTA 5751

Qy       2609 TCTATAATTGGATATAATTTAGAATGGAGTATTAATCCTTACTAAGAGGCCATATATGAA 2668

Db       5752 TCTATAATTGGATATAATTTAGAATGGAGTATTAATCCTTACTAAGAGGCCATATATGAA 5811

Qy       2669 ACTAGCCCAGACATTTATACTTTTTCAATACTTCCTTACTTTTGCTAAGCACTTCAGCAT 2728

Db       5812 ACTAGCCCAGACATTTATACTTTTTCAATACTTTCTTACTTTTGCTAAGCACTTCAGCAT 5871

Qy       2729 TTATGACTAAATATTTTGTATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGCAAA 2788

Db       5872 TTATGACTAAATATTTTGTATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGCAAA 5931

Qy       2789 ATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAACAT 2848

Db       5932 ATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAACAT 5991

Qy       2849 TTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATT 2908

Db       5992 TTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATT 6051

Qy       2909 GTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACA 2968

Db       6052 GTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACA 6111

Qy       2969 CATTCATGGCTGCTTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3028

Db       6112 CATTCATGGCTGCTTATATTCGTGAAACATATTTTATTAATTATACAGGTATTTAACTTT 6171

Qy       3029 AAAAAAAAAAAAAAAA 3044

Db       6172 CACAAGGATAATACAA 6187
```

RESULT 4
ABL10571
ID ABL10571 standard; cDNA; 2838 BP.
XX
AC ABL10571;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 26195.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB66468.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 26195; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2838 BP; 783 A; 659 C; 682 G; 714 T; 0 other;

Query Match 15.7%; Score 479.4; DB 23; Length 2838;
Best Local Similarity 52.1%; Pred. No. 1.3e-77;
Matches 1263; Conservative 0; Mismatches 1113; Indels 47; Gaps 7;

```
Qy        312 GATTACCTGAATACCCGTCCGGTTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTCCCG 371

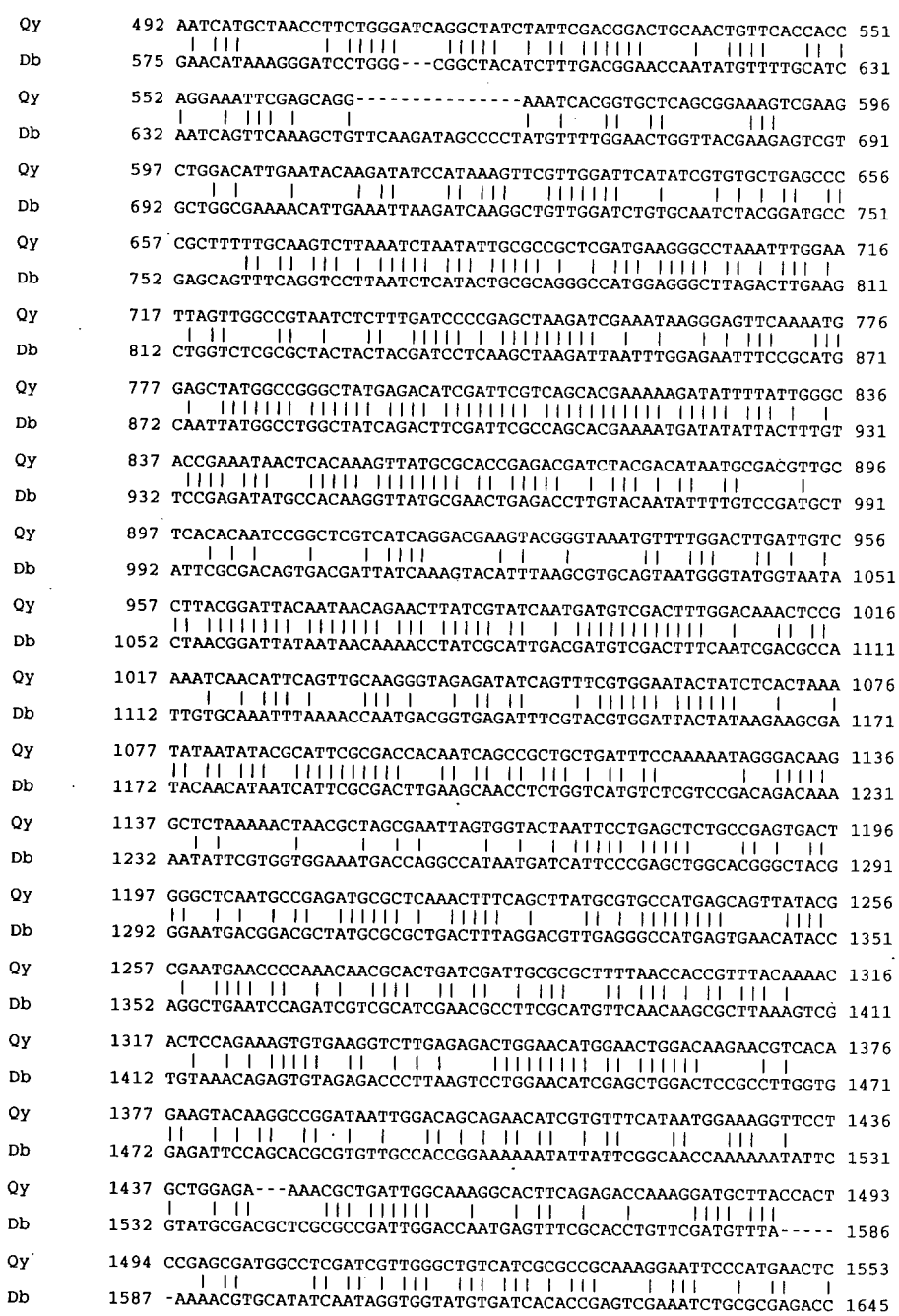
Db        395 GATTTGGTGTATTCACGTCCTCCAGGGATGACCAGCAAGAAAGGAGTGGTTGGCACTCAT 454

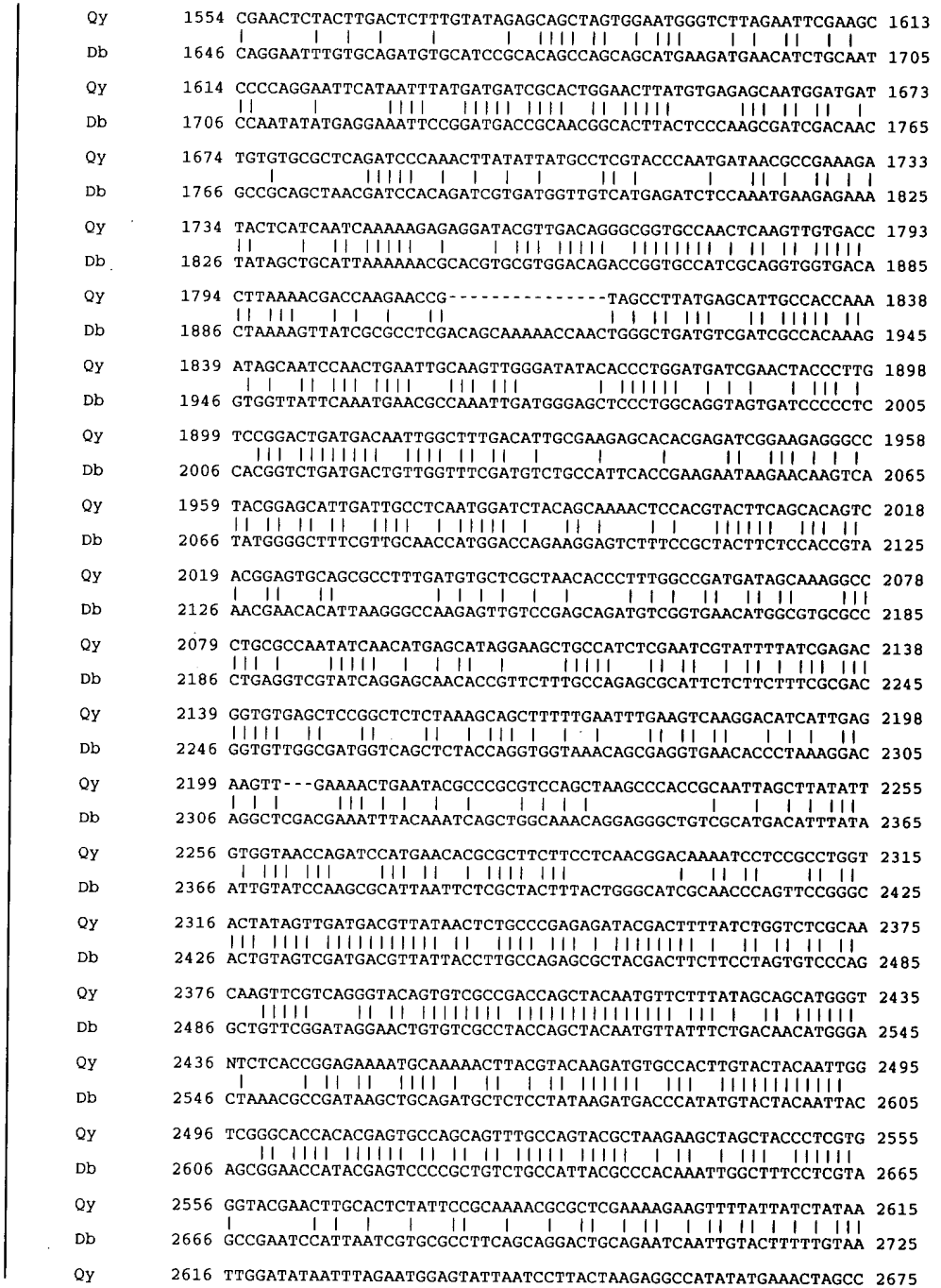
Qy        372 GTCATGCTGCAGACGAACTTTTTTCGANTAAAAACCAAGCCGGAATGGCGGATCGTTCAT 431

Db        455 ATTACCGTGCAGGCAAACTATTTTAAGGTATTAAAGCGTCCAAACTGGACCATCTACCAG 514

Qy        432 TATCACGTGGAGTTTGTGCCGACCATCGAGAATCCTCGTGTCCGTATGGGAGTTTTGTCC 491

Db        515 TACCGCGTCGATTTTACGCCTGATGTGGAGGCTACACGACTGCGACGGTCTTTCTTGTAT 574
```





```
                  ||| | ||  | |  |    | | |  |  || ||    ||||||| |  |||
Db       2726 --AACTATGACTTGCATTAAATGCGTTAACAATTTTACGATTTGATATATGTATGTAGAT 2783

Qy       2676 CAGACATTTATACTTTTTCAATA 2698
                |    | ||.    || ||||
Db       2784 ATGTACATCATTTGAATTTAATA 2806


RESULT 5
ABL08020
ID   ABL08020 standard; cDNA; 6886 BP.
XX
AC   ABL08020;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 18542.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB63917.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 18542; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 6886 BP; 1924 A; 1444 C; 1482 G; 2036 T; 0 other;

  Query Match             12.8%;  Score 389;  DB 23;  Length 6886;
  Best Local Similarity   93.1%;  Pred. No. 3.1e-61;
  Matches  407;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       2608 ATCTATAATTGGATATAATTTAGAATGGAGTATTAATCCTTACTAAGAGGCCATATATGA 2667
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATCTATAATTGGATATAATTTAGAATGGAGTATTAATCCTTACTAAGAGGCCATATATGA 60

Qy       2668 AACTAGCCCAGACATTTATACTTTTTCAATACTTCCTTACTTTTGCTAAGCACTTCAGCA 2727
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db         61 AACTAGCCCAGACATTTATACTTTTTCAATACTTTCTTACTTTTGCTAAGCACTTCAGCA 120

Qy       2728 TTTATGACTAAATATTTTGTATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGCAA 2787
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        121 TTTATGACTAAATATTTTGTATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGCAA 180

Qy       2788 AATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAACA 2847
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAACA 240

Qy       2848 TTTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTAT 2907
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTAT 300

Qy       2908 TGTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCAC 2967
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCAC 360

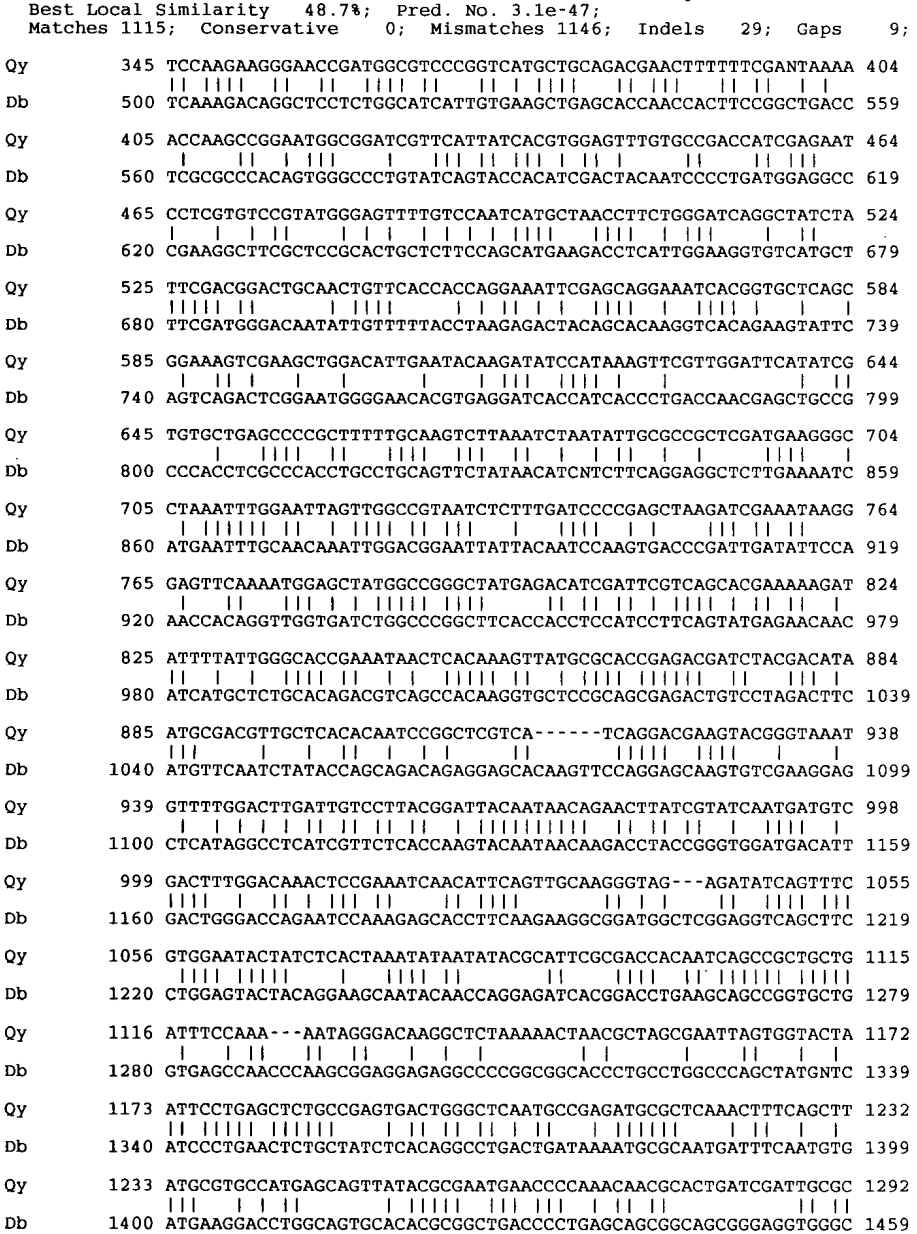
Qy       2968 ACATTCATGGCTGCTTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3027
             |||||||||||||||||||||||||| | |    |  ||  | | |   |   ||
Db        361 ACATTCATGGCTGCTTATATTCGTGAAACATATTTTATTAATTATACAGGTATTTAACTT 420

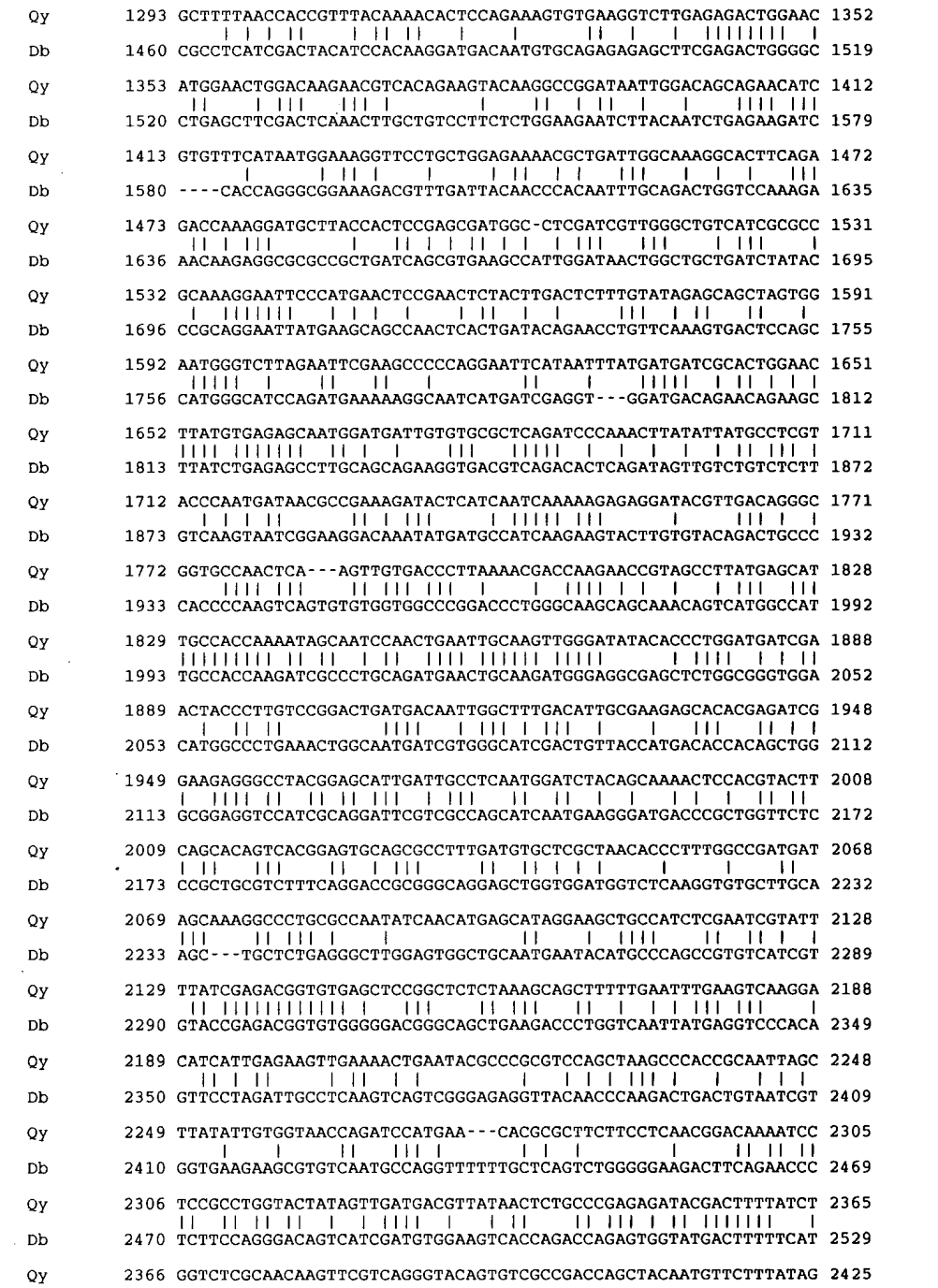
Qy       3028 AAAAAAAAAAAAAAAA 3044
                | ||  | || | ||
Db        421 TCACAAGGATAATACAA 437


RESULT 6
AAA07587
ID   AAA07587 standard; DNA; 4064 BP.
XX
AC   AAA07587;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Mouse piwi gene, designated miwi.
XX
KW   Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW   anaemia; immunodeficiency; male infertility; mouse; ds.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             191..2779
FT                   /*tag=  a
FT                   /product= miwi
FT                   /transl_except= (pos:450..452; aa:Xaa)
FT                   /transl_except= (pos:1337..1339; aa:Xaa)
FT                   /transl_except= (pos:2636..2638; aa:Xaa)
FT                   /note= "Xaa= Leu or Ile"
FT                   /transl_except= (pos:836..838; aa:Xaa)
FT                   /note= "Xaa= unspecified amino acid"
XX
PN   WO200032039-A1.
XX
PD   08-JUN-2000.
XX
PF   03-DEC-1999;   99WO-US28764.
XX
PR   04-DEC-1998;   98US-0110901.
XX
PA   (UYDU-) UNIV DUKE.
XX
PI   Lin H;
XX
DR   WPI; 2000-412085/35.
DR   P-PSDB; AAY90234.
XX
PT   Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT   therapy of diseases such as cancer and in various research and
PT   diagnostic applications -
XX
PS   Claim 19; Page 180-185; 201pp; English.
XX
CC   This sequence encodes the mouse piwi family protein, designated
CC   miwi. The piwi family nucleic acids and polypeptides are used in gene
CC   therapy of diseases such as cancer and also in various research and
```

CC diagnostic applications. The sequences can also be used to treat
CC tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ Sequence 4064 BP; 1114 A; 978 C; 1077 G; 890 T; 5 other;

Query Match 10.2%; Score 311.4; DB 21; Length 4064;
Best Local Similarity 48.7%; Pred. No. 3.1e-47;
Matches 1115; Conservative 0; Mismatches 1146; Indels 29; Gaps 9;





```
Db       2530 CGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTATGA 2589

Qy       2426 CAGCATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTA 2485

Db       2590 CAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTGTA 2649

Qy       2486 CTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGC 2545

Db       2650 CTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTGGC 2709

Qy       2546 TACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTA 2605

Db       2710 CTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTCTA 2769

Qy       2606 TTATCTATAA 2615

Db       2770 CTACCTCTAA 2779


RESULT 7
AAA07588
ID   AAA07588 standard; DNA; 3472 BP.
XX
AC   AAA07588;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Human piwi gene, designated hiwi.
XX
KW   Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW   anaemia; immunodeficiency; male infertility; human; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             167..2752
FT                   /*tag= a
FT                   /product= hiwi
FT                   /transl_except= (pos:392..394; aa:Xaa)
FT                   /transl_except= (pos:1073..1075; aa:Xaa)
FT                   /transl_except= (pos:2369..2370; aa:Xaa)
FT                   /note= "Xaa= Leu or Ile"
XX
PN   WO200032039-A1.
XX
PD   08-JUN-2000.
XX
PF   03-DEC-1999;   99WO-US28764.
XX
PR   04-DEC-1998;   98US-0110901.
XX
PA   (UYDU-) UNIV DUKE.
XX
PI   Lin H;
XX
DR   WPI; 2000-412085/35.
DR   P-PSDB; AAY90235.
XX
PT   Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT   therapy of diseases such as cancer and in various research and
PT   diagnostic applications -
XX
PS   Claim 19; Page 189-194; 201pp; English.
XX
CC   This sequence encodes the human piwi family protein, designated
CC   hiwi. The piwi family nucleic acids and polypeptides are used in gene
CC   therapy of diseases such as cancer and also in various research and
CC   diagnostic applications. The sequences can also be used to treat
CC   tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ   Sequence 3472 BP; 1044 A; 694 C; 796 G; 935 T; 3 other;
```

```
Query Match             8.3%;  Score 253;  DB 21;  Length 3472;
Best Local Similarity   47.1%;  Pred. No. 1.1e-36;
Matches 1080;  Conservative    0;  Mismatches 1184;  Indels   29;  Gaps    9;

Qy        342 GTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGANTA 401

Db        470 GAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTG 529

Qy        402 AAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAG 461

Db        530 ACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGGAA 589

Qy        462 AATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTAT 521

Db        590 GCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCAT 649

Qy        522 CTATTCGACGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTC 581

Db        650 GCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTT 709

Qy        582 AGCGGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATA 641

Db        710 TTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATGAACTT 769

Qy        642 TCGTGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCGATGAAG 701

Db        770 CCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAA 829

Qy        702 GGCCTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATA 761

Db        830 ATCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCAATTGATATT 889

Qy        762 AGGGAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAA 821

Db        890 CCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAAC 949

Qy        822 GATATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGAC 881

Db        950 AGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGAT 1009

Qy        882 ATAATGCGACGTTGCTCACACAATCCGGCTCGTCATCAGGACGAAGTACGGGT------A 935

Db       1010 TTCATGTTCAACTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAA 1069

Qy        936 AATGTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGAT 995

Db       1070 GAANTAATAGGTTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGAGTGGATGAT 1129

Qy        996 GTCGACTTTGGACAAACTCCGAAATCAACATT---CAGTTGCAAGGGTAGAGATATCAGT 1052

Db       1130 ATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGGGGTCAGC 1189

Qy       1053 TTCGTGGAATACTATCTCACTAAATATAATATACGCATTCGCGACCACAATCAGCCGCTG 1112

Db       1190 TTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTC 1249

Qy       1113 CTGATTTCCAA---AAATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTA 1169

Db       1250 TTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATG 1309

Qy       1170 CTAATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCAG 1229

Db       1310 CTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAAC 1369

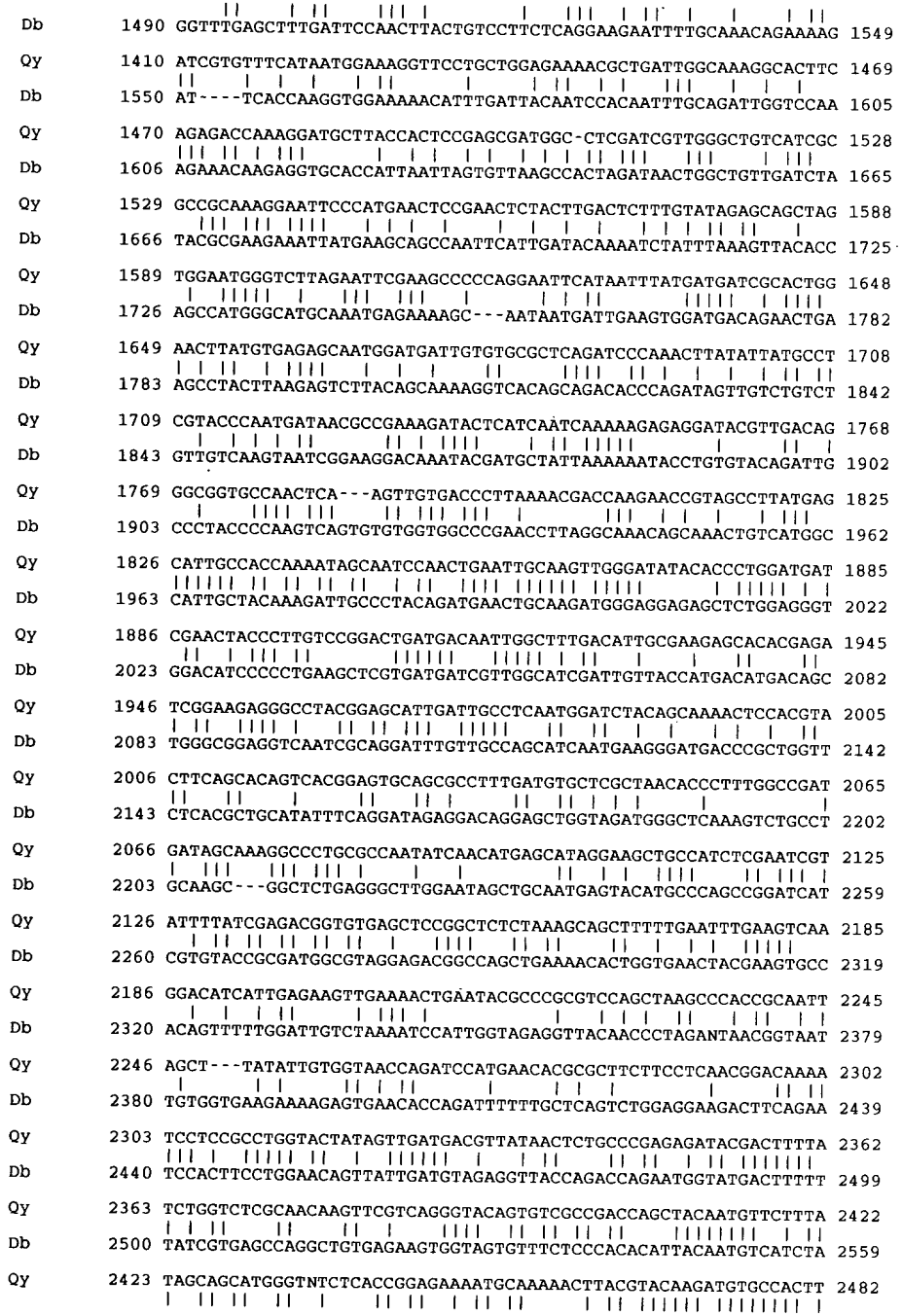
Qy       1230 CTTATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATCGATTG 1289

Db       1370 GTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTG 1429

Qy       1290 CGCGCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGACTGG 1349

Db       1430 GGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGG 1489

Qy       1350 AACATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCAGAAC 1409
```



RESULT 8
ABL10570
ID ABL10570 standard; cDNA; 5671 BP.
XX
AC ABL10570;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 26192.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB66467.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 26192; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 5671 BP; 1723 A; 1127 C; 1131 G; 1690 T; 0 other;

Query Match 7.3%; Score 221.6; DB 23; Length 5671;
Best Local Similarity 51.9%; Pred. No. 5.4e-31;
Matches 569; Conservative 0; Mismatches 520; Indels 7; Gaps 3;

```
Qy       1829 TGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGATCGA 1888
Db       3769 CGCCACAAAGGTGGTTATTCAAATGAACGCCAAATTGATGGGAGCTCCCTGGCAGGTAGT 3828

Qy       1889 ACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGATCG 1948
Db       3829 GATCCCCCTCCACGGTCTGATGACTGTTGGTTTCGATGTCTGCCATTCACCGAAGAATAA 3888

Qy       1949 GAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTT 2008
Db       3889 GAACAAGTCATATGGGGCTTTCGTTGCAACCATGGACCAGAAGGAGTCTTTCCGCTACTT 3948

Qy       2009 CAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATGAT 2068
Db       3949 CTCCACCGTAAACGAACACATTAAGGGCCAAGAGTTGTCCGAGCAGATGTCGGTGAACAT 4008

Qy       2069 AGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATT 2128
Db       4009 GGCGTGCGCCCTGAGGTCGTATCAGGAGCAACACCGTTCTTTGCCAGAGCGCATTCTCTT 4068

Qy       2129 TTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAAGGA 2188
Db       4069 CTTTCGCGACGGTGTTGGCGATGGTCAGCTCTACCAGGTGGTAAACAGCGAGGTGAACAC 4128

Qy       2189 CATCATTGAGAAGTT---GAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATT 2245
Db       4129 CCTAAAGGACAGGCTCGACGAAATTTACAAATCAGCTGGCAAACAGGAGGGCTGTCGCAT 4188

Qy       2246 AGCTTATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAATCC 2305
Db       4189 GACATTTATAATTGTATCCAAGCGCATTAATTCTCGCTACTTTACTGGGCATCGCAACCC 4248

Qy       2306 TCCGCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTATCT 2365
Db       4249 AGTTCCGGGCACTGTAGTCGATGACGTTATTACCTTGCCAGAGCGCTACGACTTCTTCCT 4308

Qy       2366 GGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAG 2425
Db       4309 AGTGTCCCAGGCTGTTCGGATAGGAACTGTGTCGCCTACCAGCTACAATGTTATTTCTGA 4368

Qy       2426 CAGCATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTA 2485
Db       4369 CAACATGGGACTAAACGCCGATAAGCTGCAGATGCTCTCCTATAAGATGACCCATATGTA 4428

Qy       2486 CTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGC 2545
Db       4429 CTACAATTACAGCGGAACCATACGAGTCCCCGCTGTCTGCCATTACGCCCACAAATTGGC 4488

Qy       2546 TACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTA 2605
Db       4489 TTTCCTCGTAGCCGAATCCATTAATCGTGCGCCTTCAGCAGGACTGCAGAATCAATTGTA 4548

Qy       2606 TTATCTATAATTGGATATAATTTAGAATGGAGTATTAATCCTTACTAAGAGGCCATATAT 2665
Db       4549 CTTTTTGTAA--AACTATGACTTGCATTAAATGCGTTAACAATTTTACGATTTGATATAT 4606

Qy       2666 GAAACTAGCCCAGACATTTATACTTTTTCAATACTTCCTTACTTTTGCTAAGCACTTCAG 2725
Db       4607 GTATGTAGATATGTACATCATTTGAATTTAATA--AAGGTATTGCACATTTATTGTCTAC 4664

Qy       2726 CATTTATGACTAAATATTTTGTATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGC 2785
Db       4665 CAATCTGAAATAATCGTTGTTTCTTTTAAATATGTAAATGTTTTTGTCTATAAAGTTCCG 4724

Qy       2786 AAAATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAA 2845
Db       4725 TTGCTGGAAAATGCATATATTCGTATGAAATGAGCTTAGTAAATAAGTTAACACAATTTA 4784

Qy       2846 CATTTAAAGTTATCTA 2861
Db       4785 CAATTAACACAATATA 4800
```

```
RESULT 9
ABL08096
ID   ABL08096 standard; cDNA; 3423 BP.
XX
AC   ABL08096;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 18770.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB63993.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 18770; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3423 BP; 1016 A; 678 C; 706 G; 1023 T; 0 other;

  Query Match             5.5%;  Score 167.2;  DB 23;  Length 3423;
  Best Local Similarity   95.0%;  Pred. No. 3.5e-21;
  Matches  172;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 CTGAGTCCAAAGCGTCGTTTTCAAAGTACTCTTTCAGTTTCCATTGTGAAGTTTTAAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2827 CTGAGTCCAAAGCGTCGTTTTCAAAGTACTCTTTCAGTTTCCATTGTGAAGTTTTAAGTG 2886

Qy         61 ATCGCGAGTGCCAAAAAGTAACAATGGCTGATGATCAGGGACGTGGACGCAGGCGTCCAN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2887 ATCGCGAGTGCCAAAAAGTAACAATGGCTGATGATCAGGGACGTGGACGCAGGCGTCCAC 2946

Qy        121 TTAACGAAGATGATTCCTCTACTTCCCGAGGTAGTGGTGATGGGCCGCGGGTGAAAGTAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||| | |
Db       2947 TTAACGAAGATGATTCCTCTACTTCCCGAGGTAGTGGTGATGGGCCGGTATGTAAACTCT 3006

Qy        181 T 181
              |
Db       3007 T 3007
```

```
RESULT 10
AAS13630
ID   AAS13630 standard; cDNA; 3649 BP.
XX
AC   AAS13630;
XX
DT   18-DEC-2001  (first entry)
XX
DE   cDNA sequence encoding mammalian Spg16.
XX
KW   Mammalian; reproductive-specific protein; male infertility; gene therapy;
KW   spermatogenesis; sperm count disorder; anti infertility;
KW   reproduction; ss.
XX
OS   Mammalia.
XX
PN   WO200166752-A2.
XX
PD   13-SEP-2001.
XX
PF   07-MAR-2001; 2001WO-US07371.
XX
PR   07-MAR-2000; 2000US-0187518.
PR   12-JAN-2001; 2001US-0261557.
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX
PI   Wang PJ,  Page DC;
XX
DR   WPI; 2001-570774/64.
DR   P-PSDB; AAU07866.
XX
PT   Novel reproduction-specific protein, useful for treating disorders of
PT   reduced sperm count, enhancing/increasing sperm count and/or sperm
PT   activity  -
XX
PS   Claim 1; Fig 15; 151pp; English.
XX
CC   The present invention relates to the isolation of novel mammalian and
CC   human reproductive-specific proteins (AAU07859-AAU07899), and the
CC   nucleic acids encoding them. The nucleic acids encoding
CC   reproductive-specific proteins are useful for diagnosing infertility
CC   which is a result of reduced sperm count, reduced sperm motility,
CC   malformed sperm or combinations of these. The sequences of the invention
CC   are useful as markers for spermatogonial cells, for identifying genes or
CC   proteins characteristic of male infertility, diagnosing or aiding in
CC   the diagnosis of infertility in men, and for contraception in which
CC   sperm production or sperm count is reduced or defective sperm is
CC   produced. Antibodies to reproductive-specific proteins are useful for
CC   determining the presence of these proteins in a sample obtained from a
CC   man being assessed for infertility, for identifying the expression of
CC   genes in particular cell type or particular developmental stage, for
CC   studies of spermatogenesis, and for immunofluorescence of germ cells or
CC   in Western blots for assessing the presence of the protein the antibody
CC   binds. The sequences of the invention are also useful for treating
CC   disorders of reduced sperm count, and for increasing sperm count and/or
CC   sperm activity. The nucleic acids of the invention are useful in gene
CC   therapy. AAS13623-AAS13647 represent cDNA sequences encoding for the
CC   mammalian reproduction-specific proteins of the present invention.
XX
SQ   Sequence 3649 BP; 947 A; 841 C; 911 G; 950 T; 0 other;

  Query Match             5.4%;  Score 164;  DB 22;  Length 3649;
  Best Local Similarity   45.8%;  Pred. No. 1.3e-20;
  Matches  849;  Conservative    0;  Mismatches  976;  Indels   30;  Gaps    7;

Qy        786 CCGGGCTATGAGACATCGATTCGTCAGCACGAAAAAGATATTTTATTGGGCACCGAAATA 845
Db          1 CCTGGCTATGCGGCTAGTATCCGGAGGACAGACGGGGGTCTCTTCCTGCTCGCTGATGTC 60

Qy        846 ACTCACAAAGTTATGCGCACCGAGACGATCTACGACATAATGCGACGT---TGCTCACAC 902
Db         61 TCTCATAAGGTCATTCGGAACGACTCTGTGCTGGATGTCATGCATGCTATCTACCAGCAG 120

Qy        903 AATCCGGCTCGTCATCAGGACGAAGTACGGGTAAATGTTTTGGACTTGATTGTCCTTACG 962
Db        121 AACAAGGAGCACTTCCAGGACGAGTGCAGCAAGCTTCTGGTTGGCAGCATTGTCATCACG 180

Qy        963 GATTACAATAACAGAACTTATCGTATCAATGATGTCGACTTTGGACAAACTCCGAAATCA 1022
Db        181 CGCTACAACAATCGTACCTACCGAATCGATGATGTGGACTGGAACAAGACCCCTAAAGAC 240

Qy       1023 ACAT---TCAGTTGCAAGGGTAGAGATATCAGTTTCGTGGAATACTATCTCACTAAATAT 1079
Db        241 AGCTTTGTCATGTCGGACGGGAAGGAAATCACATTCCTGGAATACTACAGCAAAAACTAT 300

Qy       1080 AATATACGCATTCGCGACCACAATCAGCCGCTGCTGAT------TTCCAAAAATAGGGAC 1133
Db        301 GGGATCACAGTCAAGGAAGATGACCAGCCGCTGCTGATCCACCGGCCCAGTGAGAGACAG 360

Qy       1134 AAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTACTAATTCCTGAGCTCTGCCGAGTG 1193
Db        361 AATAACCATGGCATGTTGCTGAAGGGCGAGATCCTGCTGCTGCCCGAGCTCTCCTTCATG 420

Qy       1194 ACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCAGCTTATGCGTGCCATGAGCAGTTAT 1253
Db        421 ACGGGGATCCCTGAGAAGATGAAGAAGGACTTCAGGGCCATGAAGGACTTGACTCAGCAG 480

Qy       1254 ACGCGAATGAACCCCAAACAACGCACTGATCGATTGCGCGCTTTTAACCACCGTTTACAA 1313
Db        481 ATTAACCTGAGCCCCAAGCAGCACCACGGTGCTTTGGAATGCCTGCTGCAGAGAATTTCA 540

Qy       1314 AACACTCCAGAAAGTGTGAAGGTCTTGAGAGACTGGAACATGGAACTGGACAAGAACGTC 1373
Db        541 CAAAACGAGACAGCCAGCAATGAGCTGACCCGCTGGGGGCTCAGTCTGCATAAAGATGTC 600

Qy       1374 ACAGAAGTACAAGGCCGGATAATTGGACAGCAGAACATCGTGTTTCATAATGGAAAGGTT 1433
Db        601 CACAAGATTGAAGGTCGGCTTCTGCCAATGGAGAGGATCAACTTAAGGAACACTTCATTT 660

Qy       1434 CCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTCAGAGACCAAAGGATGCTTACCACT 1493
Db        661 GTCACATCGGAGGGCCTGAACTGGGTTAAGGAAGTGACC------AGAGATGCTTCCATT 714

Qy       1494 CCGAGCGATGGCCTCGATCGTTGGGCTGTCATCGCGCCGCAAAGGAATTCCCATGAACTC 1553
Db        715 CTAACTATTCCCATGCATTTCTGGGCACTCTTTTATCCAAAGAGAGCAATGGACCAAGCC 774

Qy       1554 CGAACTCTACTTGACTCTTTGTATAGAGCAGCTAGTGGAATGGGTCTTAGAATTCGAAGC 1613
Db        775 AGAGAACTGGTTAACATGTTGGAAAAGATTGCCGGGCCCATTGGCATGCGCACAAGCCCC 834

Qy       1614 CCCCAGGAATTCATAATTTATGATGATCGCACTGGAACTTATGTGAGAGCAATGGATGAT 1673
Db        835 CCAGCCTGGGTTGAGCTGAAGGATGACCGAATAGAGACCTATATCAGGACCATTCAGTCC 894

Qy       1674 TGTGTGCGCTCAGATCCCA------AACTTATATTATGCCTCGTACCCAATGATAACGCC 1727
Db        895 TTACTGGGAGTTGAGGGGAAGATACAAATGGTCGTTTGCATCATCATGGGCACACGTGAT 954

Qy       1728 GAAAGATACTCATCAATCAAAAAGAGAGGATACGTTGACAGGGCGGTGCCAACTCAAGTT 1787
Db        955 GATCTCTATGGAGCCATCAAGAAGCTGTGCTGCGTGCAGTCCCCAGTGCCCTCACAGGTC 1014

Qy       1788 GTGACCCTTAAAACGA---CCAAGAACCGTAGCCTTATGAGCATTGCCACCAAAATAGCA 1844
Db       1015 ATCAATGTCCGAACCATTGGTCAGCCCACCAGGCTTCGGAGCGTGGCTCAGAAAATTTTA 1074

Qy       1845 ATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGATCGAACTACCCTTGTCCGGA 1904
Db       1075 CTTCAGATGAACTGTAAACTGGGTGGTGAGCTCTGGGGAGTGGATATTCCGCTGAAACAA 1134

Qy       1905 CTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGATCGGAAGAGGGCCTACGGA 1964
Db       1135 CTAATGGTGATTGGAATGGATGTGTACCATGACCCCAGCAGAGGCATGCGCTCTGTGGTC 1194
```

```
Qy       1965 GCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTTCAGCACAGTCACGGAG 2024
Db       1195 GGCTTCGTGGCCAGCATAAATCTCACACTCACCAAATGGTACTCGAGGGTGGTGTTCCAG 1254

Qy       2025 TGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATGATAGCAAAGGCCCTGCGC 2084
Db       1255 ---ATGCCACATCAGGAGATTGTGGACAGCCTGAAGCTCTGCCTGGTGGGTTCCTTGAAA 1311

Qy       2085 CAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATTTTATCGAGACGGTGTG 2144
Db       1312 AAGTATTATGAGGTGAACCATTGTCTCCCAGAGAAAATTGTGGTGTACCGAGATGGAGTG 1371

Qy       2145 AGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAAGGACATCATTGAGAAGTTG 2204
Db       1372 TCTGATGGCCAGCTAAAGACAGTTGCCAACTACGAGATCCCTCAGCTGCAGAAGTGTTTT 1431

Qy       2205 AAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATTAGCTTATATTGTGGTAACC 2264
Db       1432 GAAGCCTTTGATAACTACCACCCCAAGATGGTGGTGTTTGTAGTTCAGAAGAAAATCAGC 1491

Qy       2265 AGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAATCCTCCGCCTGGTACTATAGTT 2324
Db       1492 ACCAATCTGTACCTTGCTGCTCCTGATCACTTCGTAACCCCCTCCCCCGGGACTGTGGTT 1551

Qy       2325 GATGACGTTATAACTCTGCCCGAGAGATACGACTTTTATCTGGTCTCGCAACAAGTTCGT 2384
Db       1552 GATCATACCATAACCAGCTGTGAGTGGGTGGATTTCTACCTTCTTGCCCATCATGTGCGA 1611

Qy       2385 CAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAGCAGCATGGGTNTCTCACCG 2444
Db       1612 CAGGGCTGTGGCATACCTACACACTACATCTGTGTTCTGAACACTGCAAATCTGAGCCCT 1671

Qy       2445 GAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTACTACAATTGGTCGGGCACC 2504
Db       1672 GATCACATGCAGAGGTTGACTTTCAAACTATGCCACATGTACTGGAATTGGCCTGGTACC 1731

Qy       2505 ACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGCTACCCTCGTGGGTACGAAC 2564
Db       1732 ATCCGAGTTCCAGCTCCTTGCAAGTATGCCCACAAGCTAGCTTTCCTGTCCGGACAGATT 1791

Qy       2565 TTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTATTATCTATAATTGG 2619
Db       1792 TTGCATCATGAGCCAGCCATCCAGCTGTGTGGGAACCTGTTCTTCCTGTAACTGG 1846
```

RESULT 11
AAH15960
ID AAH15960 standard; cDNA; 2272 BP.
XX
AC AAH15960;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:14558.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 14558; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2272 BP; 609 A; 509 C; 585 G; 569 T; 0 other;

Query Match 4.1%; Score 125.8; DB 22; Length 2272;
Best Local Similarity 44.6%; Pred. No. 1e-13;
Matches 766; Conservative 0; Mismatches 923; Indels 27; Gaps 6;

```
Qy        938 TGTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGATGT 997
Db         20 TCTGGTTGGCAATATTGTTATCACCTGATATAACAATCGTACCTATCGTATTGATGATGT 79

Qy        998 CGACTTTGGACAAACTCCGAAATCAACATTC---AGTTGCAAGGGTAGAGATATCAGTTT 1054
Db         80 GGATTGGAATAAGACTCCAAAGGATAGCTTCACGATGTCTGATGGGAAAGAGATCACATT 139

Qy       1055 CGTGGAATACTATCTCACTAAATATAATATACGCATTCGCGACCACAATCAGCCGCTGCT 1114
Db        140 CTTGGAATACTACAGCAAAAATTATGGGATCACAGTTAAGGAAGAGGACCAGCCATTGCT 199

Qy       1115 GATTTCCAAAAATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTA----- 1169
Db        200 GATTCACAGGCCCAGTGAGAGACAGGATAATCATGGGATGCTGCTAAAAGGGGAAATCCT 259

Qy       1170 -CTAATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCA 1228
Db        260 GCTGCTGCCTGAGCTTTCTTTTATGACCGGAATCCCAGAGAAGATGAAGAAGGACTTCAG 319

Qy       1229 GCTTATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATCGATT 1288
Db        320 AGCCATGAAGGATTTGGCTCAGCAAATCAATCTGAGCCCCAAGCAACACCATAGTGCTTT 379

Qy       1289 GCGCGCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGACTG 1348
Db        380 GGAATGCTTGCTGCAAAGAATTGCAAAGAACGAGGCAGCCACCAATGAACTGATGCGTTG 439

Qy       1349 GAACATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCAGAA 1408
Db        440 GGGGCTCCGTCTGCAAAAGGATGTACATAAGATTGAAGGACGTGTTCTGCCAATGGAAAG 499
```

```
Qy       1409 CATCGTGTTTCATAATGGAAAGGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTT 1468
Db        500 AATTAACTTAAAAAATACTTCGTTTATCACATCTCAGGAACTAAACTGGGTTAAGGAAGT 559

Qy       1469 CAGAGACCAAAGGATGCTTACCACTCCGAGCGATGGCCTCGATCGTTGGGCTGTCATCGC 1528
Db        560 AA------CCAGAGACCCTTCCATCTTGACTATCCCCATGCATTTCTGGGCACTTTTTTA 613

Qy       1529 GCCGCAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAG 1588
Db        614 CCCAAAGAGAGCAATGGACCAGGCTCGAGAACTGGTCAACATGTTGGAGAAGATAGCCGG 673

Qy       1589 TGGAATGGGTCTTAGAATTCGAAGCCCCCAGGAATTCATAATTTATGATGATCGCACTGG 1648
Db        674 CCCCATTGGCATGCGTATGAGCCCACCGGCCTGGGTTGAACTAAAGGATGACCGAATAGA 733

Qy       1649 AACTTATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTA------TATT 1702
Db        734 GACTTATGTCAGAACCATTCAATCCACGTTAGGAGCTGAGGGGAAGATACAGATGGTTGT 793

Qy       1703 ATGCCTCGTACCCAATGATAACGCCGAAAGATACTCATCAATCAAAAAGAGAGGATACGT 1762
Db        794 TTGCATCATCATGGGCCCACGTGATGATCTCTATGGGGCCATCAAGAAGCTGTGCTGTGT 853

Qy       1763 TGACAGGGCGGTGCCAACTCAAGTTGTGACCCTTAAAACGA---CCAAGAACCGTAGCCT 1819
Db        854 GCAGTCCCCAGTGCCCTCCCAGGTTGTCAATGTTCGAACCATTGGTCAGCCCACCAGGCT 913

Qy       1820 TATGAGCATTGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTG 1879
Db        914 TCGGAGTGTGGCCCAGAAGATTTTACTTCAGATTAACTGTAAATTGGGTGGTGAGCTCTG 973

Qy       1880 GATGATCGAACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCAC 1939
Db        974 GGGAGTGGATATTCCTCTGAAACAGTTAATGGTGATCGGGATGGATGTTTACCATGACCC 1033

Qy       1940 ACGAGATCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTC 1999
Db        1034 CAGTAGAGGCATGCGCTCCGTGGTTGGCTTCGTGGCAAGCATCAATCTCACCCTCACAAA 1093

Qy       2000 CACGTACTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTG 2059
Db        1094 ATGGTATTCCCGGGTGGTGTTCCAG---ATGCCGCATCAGGAGATTGTGGACAGCCTGAA 1150

Qy       2060 GCCGATGATAGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCG 2119
Db        1151 GCTATGCCTCGTGGGCTCCTTAAAAAAGTTTTATGAGGTGAACCACTGTCTACCAGAGAA 1210

Qy       2120 AATCGTATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGA 2179
Db        1211 GATTGTGGTGTACCGTGATGGAGTGTCTGATGGCCAACTGAAGACAGTTGCCAACTATGA 1270

Qy       2180 AGTCAAGGACATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACC 2239
Db        1271 GATTCCTCAACTACAGAAGTGTTTTGAAGCTTTTGAGAATTATCAGCCCAAGATGGTGGT 1330

Qy       2240 GCAATTAGCTTATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACA 2299
Db        1331 GTTTGTAGTTCAGAAGAAAATCAGTACTAATCTATATCTGGCTGCTCCTCAGAACTTTGT 1390

Qy       2300 AAATCCTCCGCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTT 2359
Db        1391 AACTCCCACTCCTGGAACTGTGGTAGATCATACAATAACAAGCTGTGAGTGGGTGGATTT 1450

Qy       2360 TTATCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCT 2419
Db        1451 CTATCTTCTTGCCCATCATGTACGGCAGGGCTGTGGCATTCCTACGCATTATGTCTGTGT 1510

Qy       2420 TTATAGCAGCATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCA 2479
Db        1511 TCTCAACACCGCAAACCTGAGCCCTGATCATATGCAGAGGCTGACTTTCAAACTGTGCCA 1570

Qy       2480 CTTGTACTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAA 2539
```

```
Db        1571 CATGTACTGGAATTGGCCTGGCACCATCAGAGTTCCAGCTCCTTGCAAGTATGCCCACAA 1630

Qy       2540 GCTAGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAA 2599
Db        1631 GCTAGCTTTCCTGTCAGGACACATCTTGCATCATGGACCAGCCATCCAGCTGTGCGAGAA 1690

Qy       2600 GTTTTATTATCTATAATTGGATATAATTTAGAATGG 2635
Db        1691 CCTGTTCTTCCTGTGACTGCACAGCTTGGAGATGGG 1726
```

RESULT 12
AAH14032
ID AAH14032 standard; cDNA; 1733 BP.
XX
AC AAH14032;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:11143.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 11143; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification

```
CC   of the present invention.
XX
SQ   Sequence 1733 BP; 435 A; 408 C; 451 G; 439 T; 0 other;

  Query Match             3.3%;  Score 101.8;  DB 22;  Length 1733;
  Best Local Similarity   46.1%;  Pred. No. 2.2e-09;
  Matches  414;  Conservative    0;  Mismatches  478;  Indels    6;  Gaps    2;

Qy       1741 CAATCAAAAAGAGAGGATACGTTGACAGGGCGGTGCCAACTCAAGTTGTGACCCTTAAAA 1800
             | ||||| |||    | |  ||  |     | |||||  | ||  |||||  |   ||  ||
Db        293 CCATCAATAAGCTGTGCTGTGTGCAGTCCCCAGTGCCCTCCCAGGTTGTCAATGTTCGAA 352

Qy       1801 CGA---CCAAGAACCGTAGCCTTATGAGCATTGCCACCAAAATAGCAATCCAACTGAATT 1857
             | |      ||  |   || |||  |||  | |||   || ||   | | ||  | || |
Db        353 CCATTGGTCAGCCCACCAGGCTTCGGAGTGTGGCCCAGAAGATTTTACTTCAGATTAACT 412

Qy       1858 GCAAGTTGGGATATACACCCTGGATGATCGAACTACCCTTGTCCGGACTGATGACAATTG 1917
             | || |||||   |   | ||||    | ||  | ||  ||        | |||   || |
Db        413 GTAAATTGGGTGGTGAGCTCTGGGGAGTGGATATTCCTCTGAAACAGTTAATGGTGATCG 472

Qy       1918 GCTTTGACATTGCGAAGAGCACACGAGATCGGAAGAGGGCCTACGGAGCATTGATTGCCT 1977
             |  | ||  ||    |   | |  |     | | | |  ||   |  |  ||  | ||
Db        473 GGATGGATGTTTACCATGACCCCAGTAGAGGCATGCGCTCCGTGGTTGGCTTCGTGGCAA 532

Qy       1978 CAATGGATCTACAGCAAAACTCCACGTACTTCAGCACAGTCACGGAGTGCAGCGCCTTTG 2037
               ||  ||||    |  |       ||| | | |    ||     ||   |  |   |
Db        533 GCATCAATCTCACCCTCACAAAATGGTATTCCCGGGTGGTGTTCCAG---ATGCCGCATC 589

Qy       2038 ATGTGCTCGCTAACACCCTTTGGCCGATGATAGCAAAGGCCCTGCGCCAATATCAACATG 2097
             | | | | |   ||| |||   ||      | |      || |     | | | |  | |
Db        590 AGGAGATTGTGGACAGCCTGAAGCTATGCCTCGTGGGCTCCTTAAAAAAGTTTTATGAGG 649

Qy       2098 AGCATAGGAAGCTGCCATCTCGAATCGTATTTTATCGAGACGGTGTGAGCTCCGGCTCTC 2157
               | |       || |||      || ||  | || || || || |||      |||   |
Db        650 TGAACCACTGTCTACCAGAGAAGATTGTGGTGTACCGTGATGGAGTGTCTGATGGCCAAC 709

Qy       2158 TAAAGCAGCTTTTTGAATTTGAAGTCAAGGACATCATTGAGAAGTTGAAAACTGAATACG 2217
             | |||    ||    | | |||  |     |  |      ||   ||  ||  ||    |
Db        710 TGAAGACAGTTGCCAACTATGAGATTCCTCAACTACAGAAGTGTTTTGAAGCTTTTGAGA 769

Qy       2218 CCCGCGTCCAGCTAAGCCCACCGCAATTAGCTTATATTGTGGTAACCAGATCCATGAACA 2277
                  |     |     |    ||| | | |      | |  |  |     |  |
Db        770 ATTATCAGCCCAAGATGGTGGTGTTTGTAGTTCAGAAGAAAATCAGTACTAATCTATATC 829

Qy       2278 CGCGCTTCTTCCTCAACGGACAAAATCCTCCGCCTGGTACTATAGTTGATGACGTTATAA 2337
               |         |  |||      || |||  | |||||  ||| |  ||  |||  |    ||||
Db        830 TGGCTGCTCCTCAGAACTTTGTAACTCCCACTCCTGGAACTGTGGTAGATCATACAATAA 889

Qy       2338 CTCTGCCCGAGAGATACGACTTTTATCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGT 2397
             |     |||  |    || || |||||   |  | || || || || |||||    |
Db        890 CAAGCTGTGAGTGGGTGGATTTCTATCTTCTTGCCCATCATGTACGGCAGGGCTGTGGCA 949

Qy       2398 CGCCGACCAGCTACAATGTTCTTTATAGCAGCATGGGTNTCTCACCGGAGAAAATGCAAA 2457
               || ||     ||       | ||   | || |        |    || ||  | ||||| |
Db        950 TTCCTACGCATTATGTCTGTGTTCTCAACACCGCAAACCTGAGCCCTGATCATATGCAGA 1009

Qy       2458 AACTTACGTACAAGATGTGCCACTTGTACTACAATTGGTCGGGCACCACACGAGTGCCAG 2517
               || || | |||  |||||||| ||||||  ||||||  | |||||||   |||| ||||
Db       1010 GGCTGACTTTCAAACTGTGCCACATGTACTGGAATTGGCCTGGCACCATCAGAGTTCCAG 1069

Qy       2518 CAGTTTGCCAGTACGCTAAGAAGCTAGCTACCCTCGTGGGTACGAACTTGCACTCTATTC 2577
             |   |||| |||| ||  | |||||||||  |||    ||    | ||||||     |   |
Db       1070 CTCCTTGCAAGTATGCCCACAAGCTAGCTTTCCTGTCAGGACACATCTTGCATCATGAAC 1129

Qy       2578 CGCAAAACGCGCTCGAAAAGAAGTTTTATTATCTATAATTGGATATAATTTAGAATGG 2635
             |    | |  |||     ||||  | |  |  || | | || | |   |  |||  ||
Db       1130 CAGCCATCCAGCTGTGCGAGAACCTGTTCTTCCTGTGACTGCACAGCTTGGAGATGGG 1187


RESULT 13
AAS77101
ID   AAS77101 standard; cDNA; 646 BP.
XX
AC   AAS77101;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #12905.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG12914.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 12905; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 646 BP; 153 A; 160 C; 145 G; 188 T; 0 other;

  Query Match             2.7%;  Score 81.6;  DB 23;  Length 646;
  Best Local Similarity   55.5%;  Pred. No. 8.9e-06;
  Matches  156;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;

Qy       2355 GACTTTTATCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAAT 2414
               || || |||||  |  | || || || || |||||    |     || ||    ||
Db        163 GATTTCTATCTTCTTGCCCATCATGTACGGCAGGGCTGTGGCATTCCTACGCATTATGTC 222

Qy       2415 GTTCTTTATAGCAGCATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATG 2474
                 | ||   | || |       |    || ||  | |||||  |  || || | |||  ||
Db        223 TGTGTTCTCAACACCGCAAACCTGAGCCCTGATCATATGCAGAGGCTGACTTTCAAACTG 282

Qy       2475 TGCCACTTGTACTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCT 2534
              |||||| ||||||  |||||| | |||||||   |||| |||||  |||| |||| ||
Db        283 TGCCACATGTACTGGAATTGGCCTGGCACCATCAGAGTTCCAGCTCCTTGCAAGTATGCC 342
```

```
Qy       2535 AAGAAGCTAGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAA 2594
Db        343 CACAAGCTAGCTTTCCTGTCAGGACACATCTTGCATCATGAACCAGCCATCCAGCTGTGC 402

Qy       2595 AAGAAGTTTTATTATCTATAATTGGATATAATTTAGAATGG 2635
Db        403 GAGAACCTGTTCTTCCTGTGACTGCACAGCTTGGAGATGGG 443
```

RESULT 14
ABV56647/c
ID ABV56647 standard; cDNA; 449 BP.
XX
AC ABV56647;
XX
DT 17-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 56638.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 10922; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ Sequence 449 BP; 174 A; 75 C; 56 G; 144 T; 0 other;

 Query Match 2.5%; Score 77.2; DB 23; Length 449;
 Best Local Similarity 64.6%; Pred. No. 5.3e-05;
 Matches 115; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy       2870 TGTTCATTAATTTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTTTATGTAAA 2929
Db        201 TTTTTTTTTTTTTGGGGGAAAACAAATTTTTTTTTCCCTTTTTTTTTAAATTTTTTCCC 142

Qy       2930 GAAAAATTACATACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGCTTATATTC 2989
Db        141 CAAAATTTTTTTAAAAGGGGGGGGTTTTTTTTAAAAAACCCCCCTTCCCCCCTTTTTGG 82

Qy       2990 GTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3047
Db          81 GGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 24
```

RESULT 15
AAQ38808
ID AAQ38808 standard; DNA; 2750 BP.
XX
AC AAQ38808;
XX
DT 27-JUL-1993 (first entry)
XX
DE G6PD coding sequence.
XX
KW Polymerase chain reaction; glucose-6-phosphate dehydrogenase; PCR;
KW E. coli; G6PD; drug; transformation; primer; amplify; ss.
XX
OS Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT CDS 3..2258
FT /*tag= a
XX
PN WO9306125-A.
XX
PD 01-APR-1993.
XX
PF 21-SEP-1992; 92WO-US07807.
XX
PR 20-SEP-1991; 91US-0762137.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
XX
PI Kaslow DC, Shahabuddin M;
XX
DR WPI; 1993-117467/14.
DR P-PSDB; AAR33424.
XX
PT DNA segments encoding Plasmodium falciparum G6PD obtd. by PCR -
PT used to express proteins and raise antibodies for diagnosis and
PT treatment of malaria
XX
PS Claim 1; Fig 1; 36pp; English.
XX
CC This sequence encodes Plasmodium falciparum glucose-6-phosphate
CC dehydrogenase (G6PD). This sequence was isolated using the primers
CC given in AAQ38806-07. This sequence was used in the production of
CC transformed E. coli which produce a recombinant P. falciparum G6PD.
CC These transformed cells can be used in a method of screening drugs
CC for activity against P. falciparum G6PD.
XX
SQ Sequence 2750 BP; 1164 A; 260 C; 323 G; 1003 T; 0 other;

 Query Match 2.5%; Score 76.6; DB 14; Length 2750;
 Best Local Similarity 49.9%; Pred. No. 8.2e-05;
 Matches 193; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

```
Qy       2660 ATATATGAAACTAGCCCAGACATTTATACTTTTTCAATACTTCCTTACTTTTGCTAAGCA 2719
Db       2323 ATATATATATATTATATATGATTGTTTAGTATATTATTATTACCTATCTTTTATAAGATAACA 2382

Qy       2720 CTTCAGCATTTATGACTAAATATTTTGTATTTGAAATGCATTACTGCTCTTTTTTCAAAC 2779
Db       2383 TAAATGTATATATTATGACATATATATATATATATATATTATTTCACTTATCTGCCCA 2442
```

```
Qy       2780 AAAAGCAAAATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCA 2839
Db       2443 CGAACTTTATTTTTGTTTTAAAATTCTAGTATAATTAAATAAAAGAAAATATTTGGAACA 2502

Qy       2840 AACAAACATTTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTTA 2899
Db       2503 ATTTGCATTTTTTATGTATAAATAAAATTTATATAATAATATACTTTCATACTTACTTTT 2562

Qy       2900 TTTTTTATTGTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTAC 2959
Db       2563 TATTTTATTTTATTTTATTTATTTTTTAAATGTCTATTATATATACATATAAATGCGTTT 2622

Qy       2960 TTAACCACACATTCATGGCTGCTTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAA 3019
Db       2623 TCAAATAAATAATAAAAACCCATGTTTAACTAATAATATTACAAATAGAACTCAAAAAAA 2682

Qy       3020 AAAAAAAAAAAAAAAAAAAAAAAAAA 3046
Db       2683 AAAAAAATAATTATACAATGAATTAAA 2709
```

Search completed: July 21, 2003, 19:01:17
Job time : 620.788 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2003, 18:06:22 ; Search time 124.379 Seconds
(without alignments)
7512.871 Million cell updates/sec

Title: US-09-873-737A-1
Perfect score: 3047
Sequence: 1 ctgagtccaaagcgtcgttt.....aaaaaaaaaaaaaaaaaaa 3047

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	69.2	2.3	958	2	US-08-757-046A-5	Sequence 5, Appli
2	69.2	2.3	958	3	US-09-447-208-5	Sequence 5, Appli
3	69.2	2.3	958	3	US-09-135-988-5	Sequence 5, Appli
4	69.2	2.3	958	4	US-09-277-716-5	Sequence 5, Appli
5	69.2	2.3	958	4	US-08-597-274A-5	Sequence 5, Appli
6	69.2	2.3	958	4	US-08-908-909-5	Sequence 5, Appli
7	69.2	2.3	958	4	US-09-609-161B-5	Sequence 5, Appli
8	69.2	2.3	958	4	US-08-990-103-5	Sequence 5, Appli
9	68.8	2.3	1641	1	US-08-300-903A-8	Sequence 8, Appli
10	67.4	2.2	1558	1	US-08-455-550-7	Sequence 7, Appli
11	67	2.2	960	3	US-09-248-335-57	Sequence 57, Appl
12	66.2	2.2	7859	1	US-07-854-596B-4	Sequence 4, Appli
13	66.2	2.2	7859	2	US-08-450-905B-15	Sequence 15, Appl
14	66.2	2.2	7859	3	US-07-982-759F-15	Sequence 15, Appl
15	64.8	2.1	1114	4	US-09-152-060-41	Sequence 41, Appl
16	64.8	2.1	2230	4	US-08-378-313-24	Sequence 24, Appl
17	64.4	2.1	2431	1	US-07-847-743B-25	Sequence 25, Appl
18	64.4	2.1	2431	1	US-08-456-201-25	Sequence 25, Appl
19	64.4	2.1	2431	2	US-08-456-241-25	Sequence 25, Appl
20	64.4	2.1	2431	5	PCT-US92-04295A-25	Sequence 25, Appl
21	64.2	2.1	2550	6	5258287-23	Patent No. 5258287
22	64.2	2.1	3410	4	US-09-020-956-110	Sequence 110, App
23	64.2	2.1	3410	4	US-09-030-607-110	Sequence 110, App
24	64.2	2.1	3410	4	US-09-605-785-110	Sequence 110, App
25	64.2	2.1	3410	4	US-09-439-313-110	Sequence 110, App
26	64.2	2.1	3410	4	US-09-352-616A-110	Sequence 110, App
27	64.2	2.1	3410	4	US-09-602-877A-100	Sequence 100, App
28	64.2	2.1	3410	4	US-09-232-149A-110	Sequence 110, App
29	63.6	2.1	2378	4	US-08-802-805D-20	Sequence 20, Appl
30	63.2	2.1	1143	4	US-09-149-476-119	Sequence 119, App
31	63.2	2.1	1895	4	US-09-444-336-7	Sequence 7, Appli
32	63	2.1	1736	3	US-09-182-816-22	Sequence 22, Appl
c 33	63	2.1	1736	3	US-09-182-816-24	Sequence 24, Appl
34	63	2.1	1736	3	US-09-471-528-22	Sequence 22, Appl
c 35	63	2.1	1736	3	US-09-471-528-24	Sequence 24, Appl
36	63	2.1	1736	4	US-09-634-530-22	Sequence 22, Appl
c 37	63	2.1	1736	4	US-09-634-530-24	Sequence 24, Appl
38	62.6	2.1	1474	4	US-08-821-994-64	Sequence 64, Appl
39	62.6	2.1	2837	2	US-08-993-228-11	Sequence 11, Appl
40	62.4	2.0	1577	4	US-08-821-994-59	Sequence 59, Appl
c 41	62.4	2.0	5852	1	US-07-867-106-2	Sequence 2, Appli
42	62.2	2.0	1098	3	US-09-248-335-35	Sequence 35, Appl
43	62.2	2.0	1172	1	US-07-945-288-9	Sequence 9, Appli
44	62.2	2.0	1172	1	US-08-462-831-9	Sequence 9, Appli
45	62.2	2.0	1172	1	US-08-461-809-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1
US-08-757-046A-5
; Sequence 5, Application US/08757046A
; Patent No. 5876995
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; TITLE OF INVENTION: BIOLUMINESCENT ARTICLES OF MANUFACTURE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,046A
; FILING DATE: 11-25-96
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/597,274
; FILING DATE: 02-06-96
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6680-105B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 115...702
; OTHER INFORMATION: apoaequorin-encoding gene

```
;    PUBLICATION INFORMATION:
;      AUTHORS:  Inouye et al.
;      JOURNAL:  Proc. Natl. Acad. Sci. U.S.A.
;      VOLUME:  82
;      PAGES:  3154-3158
;      DATE:  (1985)
;      DOCUMENT NUMBER:  PATENT NO.:  5,093,240
US-08-757-046A-5

  Query Match             2.3%;  Score 69.2;  DB 2;  Length 958;
  Best Local Similarity   60.8%;  Pred. No. 5.8e-07;
  Matches  113;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy       2862 TCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTT 2921
              ||||   |  ||    || ||     |  ||| |  |  ||| | |||  | || || ||
Db        770 TCAATCAAAATTTTCCAAATTTTTGAACGATTTCAATCGTTTGTGTTGATTTTTGTAATT 829

Qy       2922 TATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGC 2981
                |   ||||  ||   |  |||||  ||| | |||   || | ||   |
Db        830 AGGAACAGATTAAATCGAATGATTAGTTGTTTTTTTAATCAACAGAACTTACAAATCGAA 889

Qy       2982 TTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3041
                |       ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        890 AAAGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 949

Qy       3042 AAAAAA 3047
              ||||||
Db        950 AAAAAA 955


RESULT 2
US-09-447-208-5
; Sequence 5, Application US/09447208
; Patent No. 6113886
;  GENERAL INFORMATION:
;    APPLICANT:  Bryan, Bruce
;    TITLE OF INVENTION:  BIOLUMINESCENT ARTICLES OF  MANUFACTURE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Heller Ehrman White & McAuliffe
;      STREET:  4250 Executive Square, 7th Floor
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/447,208
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  0909/135,988
;      FILING DATE:  08-17-98
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/757,046
;      FILING DATE:  11-25-96
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/597,274
;      FILING DATE:  02-06-96
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  24727-105C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-450-8400
;      TELEFAX:  619-450-8499
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  115...702
;      OTHER INFORMATION:  apoaequorin-encoding gene
;    PUBLICATION INFORMATION:
;    PUBLICATION INFORMATION:  PATENT NO.:  5,093,240
;      AUTHORS:  Inouye et al.
;      JOURNAL:  Proc. Natl. Acad. Sci. U.S.A.
;      VOLUME:  82
;      PAGES:  3154-3158
;      DATE:  (1985)
US-09-447-208-5

  Query Match             2.3%;  Score 69.2;  DB 3;  Length 958;
  Best Local Similarity   60.8%;  Pred. No. 5.8e-07;
  Matches  113;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy       2862 TCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTT 2921
              ||||   |  ||    || ||     |  ||| |  |  ||| | |||  | || || ||
Db        770 TCAATCAAAATTTTCCAAATTTTTGAACGATTTCAATCGTTTGTGTTGATTTTTGTAATT 829

Qy       2922 TATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGC 2981
                |   ||||  ||   |  |||||  ||| | |||   || | ||   |
Db        830 AGGAACAGATTAAATCGAATGATTAGTTGTTTTTTTAATCAACAGAACTTACAAATCGAA 889

Qy       2982 TTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3041
                |       ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        890 AAAGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 949

Qy       3042 AAAAAA 3047
              ||||||
Db        950 AAAAAA 955


RESULT 3
US-09-135-988-5
; Sequence 5, Application US/09135988
; Patent No. 6152358
;  GENERAL INFORMATION:
;    APPLICANT:  Bryan, Bruce
;    TITLE OF INVENTION:  BIOLUMINESCENT ARTICLES OF  MANUFACTURE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Heller Ehrman White & McAuliffe
;      STREET:  4250 Executive Square, 7th Floor
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/135,988
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/757,046
;      FILING DATE:  11-25-96
;      CLASSIFICATION:
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/597,274
;      FILING DATE:  02-06-96
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  24727-105C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-450-8400
;      TELEFAX:  619-450-8499
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  115...702
;      OTHER INFORMATION:  apoaequorin-encoding gene
;    PUBLICATION INFORMATION:
;    PUBLICATION INFORMATION:  PATENT NO.:  5,093,240
;      AUTHORS:  Inouye et al.
;      JOURNAL:  Proc. Natl. Acad. Sci. U.S.A.
;      VOLUME:  82
;      PAGES:  3154-3158
;      DATE:  (1985)
US-09-135-988-5

  Query Match             2.3%;  Score 69.2;  DB 3;  Length 958;
  Best Local Similarity   60.8%;  Pred. No. 5.8e-07;
  Matches  113;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy       2862 TCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTT 2921
              ||||   |  ||    || ||     |  ||| |  |  ||| | |||  | || || ||
Db        770 TCAATCAAAATTTTCCAAATTTTTGAACGATTTCAATCGTTTGTGTTGATTTTTGTAATT 829

Qy       2922 TATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGC 2981
                 |   ||||   ||   |  |||||  ||| | |||   || | ||   |
Db        830 AGGAACAGATTAAATCGAATGATTAGTTGTTTTTTTAATCAACAGAACTTACAAATCGAA 889

Qy       2982 TTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3041
                |        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        890 AAAGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 949

Qy       3042 AAAAAA 3047
              ||||||
Db        950 AAAAAA 955


RESULT 4
US-09-277-716-5
; Sequence 5, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
;  APPLICANT: Bryan, Bruce
;  APPLICANT: Szent-Gyorgyi, Christopher
;  APPLICANT: PROLUME, LTD.
;  TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
;  CURRENT APPLICATION NUMBER: US/09/277,716A
;  CURRENT FILING DATE: 1999-03-26
;  EARLIER APPLICATION NUMBER: 60/102,939
;  EARLIER FILING DATE: 1998-10-01
;  EARLIER APPLICATION NUMBER: 60/089,367
;  EARLIER FILING DATE: 1998-06-15
;  EARLIER APPLICATION NUMBER: 60/079,624
;  EARLIER FILING DATE: 1998-03-27
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 958
;   TYPE: DNA
;   ORGANISM: Aequorea (luminescent jellyfish)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (115)..(702)
;   FEATURE:
;   OTHER INFORMATION: Apoaequorin-encoding gene
;  PUBLICATION INFORMATION:
;   PATENT DOCUMENT NUMBER: 5,093,240
;   PATENT FILING DATE: 1987-10-08
;   PUBLICATION DATE: 1992-03-03
;  PUBLICATION INFORMATION:
;   AUTHORS: Inouye, S.
;   TITLE: Cloning and sequence analysis of cDNA for the luminescent protein    aequo
;   JOURNAL: Proc. Natl. Acad. Sci. USA
;   VOLUME: 82(10)
;   PAGES: 3154-3158
;   DATE: 1985-05
US-09-277-716-5

  Query Match             2.3%;  Score 69.2;  DB 4;  Length 958;
  Best Local Similarity   60.8%;  Pred. No. 5.8e-07;
  Matches  113;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy       2862 TCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTT 2921
              ||||   |  ||    || ||     |  ||| |  |  ||| | |||  | || || ||
Db        770 TCAATCAAAATTTTCCAAATTTTTGAACGATTTCAATCGTTTGTGTTGATTTTTGTAATT 829

Qy       2922 TATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGC 2981
                 |   ||||   ||   |  |||||  ||| | |||   || | ||   |
Db        830 AGGAACAGATTAAATCGAATGATTAGTTGTTTTTTTAATCAACAGAACTTACAAATCGAA 889

Qy       2982 TTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3041
                |        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        890 AAAGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 949

Qy       3042 AAAAAA 3047
              ||||||
Db        950 AAAAAA 955


RESULT 5
US-08-597-274A-5
; Sequence 5, Application US/08597274A
; Patent No. 6247995
;  GENERAL INFORMATION:
;    APPLICANT:  Bryan, Bruce
;    TITLE OF INVENTION:  BIOLUMINESCENT NOVELTY ITEMS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Martin, Haller & McClain
;      STREET:  1660 Union Street
;      CITY:  San Diego
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92101-2926
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/597,274A
;      FILING DATE:  02/06/96
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
```

```
;       NAME:  Seidman, Stephanie L
;       REGISTRATION NUMBER:  33,779
;       REFERENCE/DOCKET NUMBER:  6680-105
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  619-238-0999
;       TELEFAX:  619-238-0062
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  958 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FRAGMENT TYPE:
;     ORIGINAL SOURCE:
;     FEATURE:
;       NAME/KEY:  Coding Sequence
;       LOCATION:  115...702
;       OTHER INFORMATION:  apoaequorin-encoding gene
;     PUBLICATION INFORMATION:
;       DOCUMENT NUMBER:  5,093,240
;       AUTHORS:  Inouye et al.
;       JOURNAL:  Proc. Natl. Acad. Sci. U.S.A.
;       VOLUME:  82
;       PAGES:  3154-3158
;       DATE:  (1985)
US-08-597-274A-5

  Query Match             2.3%;  Score 69.2;  DB 4;  Length 958;
  Best Local Similarity   60.8%;  Pred. No. 5.8e-07;
  Matches  113;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy       2862 TCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTT 2921
             ||||   |  ||    || ||     |  ||| |  |  ||| | |||  | || || ||
Db        770 TCAATCAAAATTTTCCAAATTTTTGAACGATTTCAATCGTTTGTGTTGATTTTTGTAATT 829

Qy       2922 TATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGC 2981
                 |    ||||   ||   |   |||||  ||| | |||   || | ||   |
Db        830 AGGAACAGATTAAATCGAATGATTAGTTGTTTTTTTAATCAACAGAACTTACAAATCGAA 889

Qy       2982 TTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3041
                |        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        890 AAAGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 949

Qy       3042 AAAAAA 3047
              ||||||
Db        950 AAAAAA 955


RESULT 6
US-08-908-909-5
; Sequence 5, Application US/08908909
; Patent No. 6416960
;  GENERAL INFORMATION:
;    APPLICANT:  Bryan, Bruce
;    TITLE OF INVENTION:  DETECTION AND VISUALIZATION OF
;    TITLE OF INVENTION:  NEOPLASTIC TISSUES AND OTHER TISSUES
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Martin, Haller & McClain
;      STREET:  1660 Union Street
;      CITY:  San Diego
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92101-2926
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/908,909
;      FILING DATE:  08-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/023,374
;      FILING DATE:  08-AUG-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  6680-108
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-238-0999
;      TELEFAX:  619-238-0062
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  115...702
;      OTHER INFORMATION:  apoaequorin-encoding gene
;    PUBLICATION INFORMATION:
;      AUTHORS:  Inouye et al.
;      JOURNAL:  Proc. Natl. Acad. Sci. U.S.A.
;      VOLUME:  82
;      PAGES:  3154-3158
;      DATE:  (1985)
;      DOCUMENT NUMBER:  5,093,240
US-08-908-909-5

  Query Match             2.3%;  Score 69.2;  DB 4;  Length 958;
  Best Local Similarity   60.8%;  Pred. No. 5.8e-07;
  Matches  113;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy       2862 TCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTT 2921
             ||||   |  ||    || ||     |  ||| |  |  ||| | |||  | || || ||
Db        770 TCAATCAAAATTTTCCAAATTTTTGAACGATTTCAATCGTTTGTGTTGATTTTTGTAATT 829

Qy       2922 TATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGC 2981
                 |    ||||   ||   |   |||||  ||| | |||   || | ||   |
Db        830 AGGAACAGATTAAATCGAATGATTAGTTGTTTTTTTAATCAACAGAACTTACAAATCGAA 889

Qy       2982 TTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3041
                |        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        890 AAAGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 949

Qy       3042 AAAAAA 3047
              ||||||
Db        950 AAAAAA 955


RESULT 7
US-09-609-161B-5
; Sequence 5, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
;  APPLICANT: Bryan, Bruce
;  APPLICANT:  Szent-Gyorgyi, Christopher
;  APPLICANT:  PROLUME, LTD.
;  TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
;  TITLE OF INVENTION:  AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS,
;  TITLE OF INVENTION:  SCREENING AND NOVELTY ITEMS
;  FILE REFERENCE: 24729-121B
```

```
;  CURRENT APPLICATION NUMBER: US/09/609,161B
;  CURRENT FILING DATE:  2000-06-30
;  PRIOR APPLICATION NUMBER: 09/277,716
;  PRIOR FILING DATE: 1999-03-26
;  PRIOR APPLICATION NUMBER: 60/102,939
;  PRIOR FILING DATE: 1998-10-01
;  PRIOR APPLICATION NUMBER: 60/089,367
;  PRIOR FILING DATE: 1998-06-15
;  PRIOR APPLICATION NUMBER: 60/079,624
;  PRIOR FILING DATE: 1998-03-27
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 958
;   TYPE: DNA
;   ORGANISM: Aequorea (luminescent jellyfish)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (115)..(702)
;   OTHER INFORMATION: Apoaequorin-encoding gene
;  PUBLICATION INFORMATION:
;   PATENT DOCUMENT NUMBER: US 07/105,602
;   PATENT FILING DATE: 1987-10-08
;   PUBLICATION DATE: 1992-03-03
;  PUBLICATION INFORMATION:
;   AUTHORS: Inouye, S.
;   AUTHORS: No. 6436682uchi, M.
;   AUTHORS: Sakaki, Y.
;   AUTHORS: Takagi, Y.
;   AUTHORS: Miyata, T.
;   AUTHORS: Iwanaga, S.
;   AUTHORS: Miyata, T.
;   AUTHORS: Tsuji, F.I.
;   TITLE: Cloning and sequence analysis of cDNA for the luminescent protein
;   TITLE: aequorin
;   JOURNAL: Proc. Natl. Acad. Sci. USA
;   VOLUME: 82(10)
;   PAGES: 3154-3158
;   DATE: 1985-05
US-09-609-161B-5

  Query Match             2.3%;  Score 69.2;  DB 4;  Length 958;
  Best Local Similarity   60.8%;  Pred. No. 5.8e-07;
  Matches  113;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy       2862 TCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTT 2921
             ||||   |  ||    || ||     |  ||| |  |  ||| | |||  | || || ||
Db        770 TCAATCAAAATTTTCCAAATTTTTGAACGATTTCAATCGTTTGTGTTGATTTTTGTAATT 829

Qy       2922 TATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGC 2981
                |    ||||   ||   |   |||||  ||| | |||   || | ||  |
Db        830 AGGAACAGATTAAATCGAATGATTAGTTGTTTTTTTAATCAACAGAACTTACAAATCGAA 889

Qy       2982 TTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3041
                |        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        890 AAAGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 949

Qy       3042 AAAAAA 3047
              ||||||
Db        950 AAAAAA 955


RESULT 8
US-08-990-103-5
; Sequence 5, Application US/08990103
; Patent No. 6458547
;  GENERAL INFORMATION:
;    APPLICANT:  Bruce J. Bryan
;    APPLICANT:  Stephen Gaalema
;    APPLICANT:  Randall B. Murphy
;    TITLE OF INVENTION:  APPARATUS AND METHOD FOR DETECTING AND
;    TITLE OF INVENTION:  IDENTIFYING INFECTIOUS AGENTS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Martin, Haller & McClain
;      STREET:  1660 Union Street
;      CITY:  San Diego
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92101-2926
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/990,103
;      FILING DATE:  12-DEC-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/037,675,
;      FILING DATE:  02-FEB-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/033,745
;      FILING DATE:  12-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  6680-112
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-238-0999
;      TELEFAX:  619-238-0062
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  115...702
;      OTHER INFORMATION:  apoaequorin-encoding gene
;    PUBLICATION INFORMATION:
;      AUTHORS:  Inouye et al.
;      JOURNAL:  Proc. Natl. Acad. Sci. U.S.A.
;      VOLUME:  82
;      PAGES:  3154-3158
;      DATE:  (1985)
;      DOCUMENT NUMBER:  5,093,240
US-08-990-103-5

  Query Match             2.3%;  Score 69.2;  DB 4;  Length 958;
  Best Local Similarity   60.8%;  Pred. No. 5.8e-07;
  Matches  113;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy       2862 TCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTT 2921
             ||||   |  ||    || ||     |  ||| |  |  ||| | |||  | || || ||
Db        770 TCAATCAAAATTTTCCAAATTTTTGAACGATTTCAATCGTTTGTGTTGATTTTTGTAATT 829

Qy       2922 TATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGC 2981
                |    ||||   ||   |   |||||  ||| | |||   || | ||  |
Db        830 AGGAACAGATTAAATCGAATGATTAGTTGTTTTTTTAATCAACAGAACTTACAAATCGAA 889

Qy       2982 TTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3041
                |        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        890 AAAGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 949

Qy       3042 AAAAAA 3047
              ||||||
Db        950 AAAAAA 955
```

```
RESULT 9
US-08-300-903A-8
; Sequence 8, Application US/08300903A
; Patent No. 5591630
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, Dirk M
;    APPLICANT:  Giri, Judith G
;    TITLE OF INVENTION:  Interleukin-15 Receptors
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Immunex Corporation
;      STREET:  51 University Street
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Apple Operating System 7.1
;      SOFTWARE:  Microsoft Word for Apple, Version 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/300,903A
;      FILING DATE:  06-SEPTEMBER-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  USSN 08/236,919
;      FILING DATE:  06-MAY-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perkins, Patricia Anne
;      REGISTRATION NUMBER:  34,695
;      REFERENCE/DOCKET NUMBER:  2822-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-587-0430
;      TELEFAX:  206-233-0644
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1641 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..839
US-08-300-903A-8

  Query Match             2.3%;  Score 68.8;  DB 1;  Length 1641;
  Best Local Similarity   56.0%;  Pred. No. 8.6e-07;
  Matches  130;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy       2814 AAGCATAAGACCAGAGGAAATTCCCAAACAAACATTTAAAGTTATCTATCAAGACATGTT 2873
             ||| | || ||   |   | | ||     |   || ||    | | |||  | || | | |
Db       1410 AAGAAGAAAACTCAACTCAGTGCCATTTTACGAATATATGCGTTTATATTTATACTTCCT 1469

Qy       2874 CATTAATTTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTTTATGTAAAGAAA 2933
               |  |||     ||||  |  | |||| ||| |  | |    | ||   |   |  ||
Db       1470 TGTCTATTATATCTATACATTATATATTATTTGTATTTTGACATTGTACCTTGTATAAAC 1529

Qy       2934 AATTACATACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGCTTATATTCGTGA 2993
               ||    | ||||| || ||   | |   ||  | | |   |          | |   |
Db       1530 AAAATAAAACATCTATTTTCAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1589

Qy       2994 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3045
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1590 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1641
```

```
RESULT 10
US-08-455-550-7
; Sequence 7, Application US/08455550
; Patent No. 5670338
;  GENERAL INFORMATION:
;    APPLICANT:  MURAKAMI, KAZUO
;    APPLICANT:  UENO, NAOTO
;    APPLICANT:  KATO, YUKIO
;    TITLE OF INVENTION:  XENOPUS LAEVIS BONE MORPHOGENETIC PROTEINS AND USE THE
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dike, Bronstein, Roberts & Cushman
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/455,550
;      FILING DATE:  31-MAY-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/056,564
;      FILING DATE:  30-APR-1993
;      APPLICATION NUMBER:  07/577,892
;      FILING DATE:  05-SEP-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Eisenstein, Ronald I
;      REGISTRATION NUMBER:  30628
;      REFERENCE/DOCKET NUMBER:  40302-FWC-DIV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;      TELEX:  200291
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1558 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-455-550-7

  Query Match             2.2%;  Score 67.4;  DB 1;  Length 1558;
  Best Local Similarity   60.8%;  Pred. No. 1.8e-06;
  Matches  110;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy       2867 ACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTTTATGT 2926
              || |   | ||| ||  || | | ||    |  || |||| |||   || | |   ||| |
Db       1367 ACCTTGACCTTATTTATGACTTTTATGTGTAAATGTTTTTTTGACAATATGATCATATAT 1426

Qy       2927 AAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGCTTATA 2986
                || ||| || ||  ||   |   |  | |  ||  | | |          |   |
Db       1427 TTTGACAAAATATATTTATAACTACGTATTAAAAGAAAAAAAAAAAATAAAATAAGTCAT 1486

Qy       2987 TTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3046
              |   | ||| | ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TATTTTAAACATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1546

Qy       3047 A 3047
              |
Db       1547 A 1547
```

```
RESULT 11
US-09-248-335-57
; Sequence 57, Application US/09248335
; Patent No. 6096504
; GENERAL INFORMATION:
;  APPLICANT: MCGONIGLE, BRIAN
;  APPLICANT: O'KEEF, DANIEL
;  TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
;  FILE REFERENCE: CL-1128-A
;  CURRENT APPLICATION NUMBER: US/09/248,335
;  CURRENT FILING DATE: 1999-02-10
;  EARLIER APPLICATION NUMBER: 08/924,759
;  EARLIER FILING DATE: 1997-September-05
;  NUMBER OF SEQ ID NOS: 74
;  SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 57
;   LENGTH: 960
;   TYPE: DNA
;   ORGANISM: maize
US-09-248-335-57

  Query Match             2.2%;  Score 67;  DB 3;  Length 960;
  Best Local Similarity   62.0%;  Pred. No. 1.9e-06;
  Matches  106;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy       2877 TAATTTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTTTATGTAAAGAAAAAT 2936
             ||||   ||||| |||||   | |  |||| |  | || |  ||      |  || ||
Db        775 TAATAAAGAATACAATTAGGGTGTCGTTTAGTTCACATGTCGGTAACGTAATGGATAACC 834

Qy       2937 TACATACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGCTTATATTCGTGAAAA 2996
              | |    |  ||  |   |    |||   ||     |      || ||| |    ||||
Db        835 GATAACATTAAATCATGTTTGTTATAAGTCCAATCGTAATCGATCTCATACTACAAAAAA 894

Qy       2997 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3047
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        895 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 945


RESULT 12
US-07-854-596B-4
; Sequence 4, Application US/07854596B
; Patent No. 5434073
;  GENERAL INFORMATION:
;    APPLICANT:  Dawson, Keith M
;    APPLICANT:  Hunter, Michael G
;    APPLICANT:  Czaplewski, Lloyd G
;    TITLE OF INVENTION:  Proteins and nucleic acids
;    NUMBER OF SEQUENCES:  73
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dr. John J. McDonnell
;      STREET:  Ten South Wacker Drive, Suite 3000
;      CITY:  Chicago
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/854,596B
;      FILING DATE:  03-JUN-1992
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McDonnell, John J
;      REGISTRATION NUMBER:  26,949
;      REFERENCE/DOCKET NUMBER:  92,337
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-715-1000
;      TELEFAX:  312-715-1234
;      TELEX:  910-221-5317
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7859 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..7859
;      OTHER INFORMATION:  /note= "sequence of plasmid pSW6"
US-07-854-596B-4

  Query Match             2.2%;  Score 66.2;  DB 1;  Length 7859;
  Best Local Similarity   81.1%;  Pred. No. 5.6e-06;
  Matches   77;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy       2953 TGCTTACTTAACCACACATTCATGGCTGCTTATATTCGTGAAAAAAAAAAAAAAAAAAAA 3012
              | ||| |||||  | |   || |   |  | ||||| ||  ||||||||||||||||||||
Db       5569 TCCTTGCTTAAAAAGATTCTCTTTTTTTATGATATTTGTACAAAAAAAAAAAAAAAAAAAA 5628

Qy       3013 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3047
              ||||||||||||||||||||||||||||||||||||
Db       5629 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5663


RESULT 13
US-08-450-905B-15
; Sequence 15, Application US/08450905B
; Patent No. 5856301
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Stem Cell Inhibiting Proteins
;    NUMBER OF SEQUENCES:  178
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HALE and DORR
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  MA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/450,905B
;      FILING DATE:  26-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/982,759
;      FILING DATE:  08-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9127319.3
;      FILING DATE:  23-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9221587.0
;      FILING DATE:  14-OCT-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BAKER, HOLLIE L.
;      REGISTRATION NUMBER:  31,321
;      REFERENCE/DOCKET NUMBER:  102.378.120DV-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-526-6110
;      TELEFAX:  617-526-5000
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7859 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA
US-08-450-905B-15

  Query Match             2.2%;  Score 66.2;  DB 2;  Length 7859;
```

```
  Best Local Similarity   81.1%;  Pred. No. 5.6e-06;
  Matches   77;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy       2953 TGCTTACTTAACCACACATTCATGGCTGCTTATATTCGTGAAAAAAAAAAAAAAAAAAAA 3012
              | ||| |||||  | |   || |   |  | ||||| ||  ||||||||||||||||||
Db       5569 TCCTTGCTTAAAAAGATTCTCTTTTTTTATGATATTTGTACAAAAAAAAAAAAAAAAAAAA 5628

Qy       3013 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3047
              ||||||||||||||||||||||||||||||||||||
Db       5629 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5663


RESULT 14
US-07-982-759F-15
; Sequence 15, Application US/07982759F
; Patent No. 6057123
;  GENERAL INFORMATION:
;    APPLICANT:  CRAIG, Stewart
;    APPLICANT:  GEORGE, Michael
;    APPLICANT:  EDWARDS, Richard Mark
;    APPLICANT:  CZAPLEWSKI, Lloyd George
;    APPLICANT:  GILBERT, Richard
;    TITLE OF INVENTION:  Stem Cell Inhibiting Proteins
;    NUMBER OF SEQUENCES:  178
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HALE and DORR LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  MA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/982,759F
;      FILING DATE:  08-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9127319.3
;      FILING DATE:  23-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9221587.0
;      FILING DATE:  14-OCT-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BAKER, HOLLIE L.
;      REGISTRATION NUMBER:  31,321
;      REFERENCE/DOCKET NUMBER:  102378.120
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-526-6000
;      TELEFAX:  617-526-5000
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7859 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA
US-07-982-759F-15

  Query Match             2.2%;  Score 66.2;  DB 3;  Length 7859;
  Best Local Similarity   81.1%;  Pred. No. 5.6e-06;
  Matches   77;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy       2953 TGCTTACTTAACCACACATTCATGGCTGCTTATATTCGTGAAAAAAAAAAAAAAAAAAAA 3012
              | ||| |||||  | |   || |   |  | ||||| ||  ||||||||||||||||||
Db       5569 TCCTTGCTTAAAAAGATTCTCTTTTTTTATGATATTTGTACAAAAAAAAAAAAAAAAAAAA 5628

Qy       3013 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3047
              ||||||||||||||||||||||||||||||||||||
Db       5629 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5663


RESULT 15
US-09-152-060-41
; Sequence 41, Application US/09152060
; Patent No. 6448230
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 28 Human Secreted Proteins
;  FILE REFERENCE: PZ003P1.US
;  CURRENT APPLICATION NUMBER: US/09/152,060
;  CURRENT FILING DATE: 1998-09-11
;  EARLIER APPLICATION NUMBER: PCT/US98/04858
;  EARLIER FILING DATE: 1998-03-12
;  EARLIER APPLICATION NUMBER: 60/040,762
;  EARLIER FILING DATE: 1997-03-14
;  EARLIER APPLICATION NUMBER: 60/040,710
;  EARLIER FILING DATE: 1997-03-14
;  EARLIER APPLICATION NUMBER: 60/050,934
;  EARLIER FILING DATE: 1997-05-30
;  EARLIER APPLICATION NUMBER: 60/048,100
;  EARLIER FILING DATE: 1997-05-30
;  EARLIER APPLICATION NUMBER: 60/048,357
;  EARLIER FILING DATE: 1997-05-30
;  EARLIER APPLICATION NUMBER: 60/048,189
;  EARLIER FILING DATE: 1997-05-30
;  EARLIER APPLICATION NUMBER: 60/057,765
;  EARLIER FILING DATE: 1997-09-05
;  EARLIER APPLICATION NUMBER: 60/048,970
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/068,368
;  EARLIER FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 118
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
;  LENGTH: 1114
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-152-060-41

  Query Match             2.1%;  Score 64.8;  DB 4;  Length 1114;
  Best Local Similarity   68.9%;  Pred. No. 6.3e-06;
  Matches   84;  Conservative    3;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       2926 TAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGCTTAT 2985
              |||  || ||| |         | ||  |    ||     ||   | :  |||||: |||:
Db        969 TAAGAAACAATAATTATGGCACCTGAATTAGGACAGTGACATTAAAKGTTGGCTKTTTAW 1028

Qy       2986 ATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3045
              |||    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1029 ATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1088

Qy       3046 AA 3047
              ||
Db       1089 AA 1090


Search completed: July 22, 2003, 06:25:06
Job time : 126.379 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2003, 18:26:42 ; Search time 588.209 Seconds
(without alignments)
10686.606 Million cell updates/sec

Title: US-09-873-737A-1
Perfect score: 3047
Sequence: 1 ctgagtccaaagcgtcgttt.....aaaaaaaaaaaaaaaaaaa 3047

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
11: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq3:*
13: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	3044	99.9	3047	10	US-09-873-737A-1	Sequence 1, Appli
	2	311.4	10.2	4064	10	US-09-873-737A-3	Sequence 3, Appli
	3	256.6	8.4	2328	14	US-10-043-774B-1	Sequence 1, Appli
	4	253	8.3	3472	10	US-09-873-737A-5	Sequence 5, Appli
	5	164	5.4	3649	10	US-09-801-574-15	Sequence 15, Appl
	6	161.6	5.3	2292	11	US-09-998-598-2602	Sequence 2602, Ap
	7	149	4.9	2009	15	US-10-029-495-8	Sequence 8, Appli
c	8	81.2	2.7	393	11	US-09-960-352-4582	Sequence 4582, Ap
c	9	77	2.5	397	11	US-09-960-352-13784	Sequence 13784, A
c	10	76.2	2.5	419	11	US-09-960-352-11234	Sequence 11234, A
c	11	75.2	2.5	375	11	US-09-960-352-15014	Sequence 15014, A
c	12	74.2	2.4	337	11	US-09-960-352-6976	Sequence 6976, Ap
c	13	73.6	2.4	638	15	US-10-198-846-8560	Sequence 8560, Ap
	14	73.4	2.4	1046	10	US-09-925-297-307	Sequence 307, App
c	15	72.8	2.4	509	11	US-09-834-975-504	Sequence 504, App
	16	72.8	2.4	981	10	US-09-780-717-4	Sequence 4, Appli
c	17	72.6	2.4	6327	15	US-10-239-676-151	Sequence 151, App
	18	72.4	2.4	673	15	US-10-106-698-1147	Sequence 1147, Ap
c	19	72.4	2.4	859	15	US-10-198-846-6721	Sequence 6721, Ap
	20	72.2	2.4	250	11	US-09-764-846-134	Sequence 134, App
	21	72.2	2.4	250	15	US-10-091-483-134	Sequence 134, App
c	22	72.2	2.4	1194	15	US-10-106-698-1692	Sequence 1692, Ap
	23	72.2	2.4	1336	11	US-09-764-846-62	Sequence 62, Appl
	24	72.2	2.4	1336	15	US-10-091-483-62	Sequence 62, Appl
	25	72.2	2.4	10365	15	US-10-098-939-1	Sequence 1, Appli
	26	72.2	2.4	10381	15	US-10-098-939-2	Sequence 2, Appli
	27	72.2	2.4	10423	15	US-10-098-939-3	Sequence 3, Appli
c	28	71.4	2.3	858	15	US-10-198-846-7035	Sequence 7035, Ap
c	29	71	2.3	277	11	US-09-960-352-12673	Sequence 12673, A
c	30	71	2.3	17848	15	US-10-239-676-28	Sequence 28, Appl
c	31	70.6	2.3	892	15	US-10-198-846-7154	Sequence 7154, Ap
	32	69.8	2.3	497	12	US-09-918-995-12519	Sequence 12519, A
	33	69.4	2.3	395	12	US-09-918-995-3637	Sequence 3637, Ap
	34	69.2	2.3	958	10	US-09-803-211-5	Sequence 5, Appli
	35	69.2	2.3	958	10	US-09-746-485A-5	Sequence 5, Appli
	36	69.2	2.3	958	12	US-09-808-898-5	Sequence 5, Appli
	37	69.2	2.3	958	15	US-10-126-139-5	Sequence 5, Appli
	38	69.2	2.3	958	15	US-10-126-798-5	Sequence 5, Appli
	39	69.2	2.3	958	15	US-10-126-777-5	Sequence 5, Appli
	40	69.2	2.3	1484	11	US-09-925-300-248	Sequence 248, App
c	41	69	2.3	239	11	US-09-960-352-11438	Sequence 11438, A
	42	69	2.3	329	11	US-09-880-107-706	Sequence 706, App
c	43	69	2.3	397	15	US-10-198-846-1348	Sequence 1348, Ap
	44	69	2.3	3413	10	US-09-796-858-1	Sequence 1, Appli
	45	69	2.3	3413	10	US-09-796-858-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-873-737A-1
; Sequence 1, Application US/09873737A
; Patent No. US20020076797A1
; GENERAL INFORMATION:
; APPLICANT: Duke University
; APPLICANT: Lin, Haifan
; TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
; TITLE OF INVENTION: PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
; FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2
; CURRENT APPLICATION NUMBER: US/09/873,737A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/US99/28764
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/110,901
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3047
; TYPE: DNA
; ORGANISM: Drosophila sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (84)..(2612)
; NAME/KEY: misc_feature
; LOCATION: (120)
; OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (399)
; OTHER INFORMATION: n=a or t, Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (2436)
; OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
US-09-873-737A-1

Query Match 99.9%; Score 3044; DB 10; Length 3047;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3047; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CTGAGTCCAAAGCGTCGTTTTCAAAGTACTCTTTCAGTTTCCATTGTGAAGTTTTAAGTG 60
Db       1 CTGAGTCCAAAGCGTCGTTTTCAAAGTACTCTTTCAGTTTCCATTGTGAAGTTTTAAGTG 60

Qy      61 ATCGCGAGTGCCAAAAAGTAACAATGGCTGATGATCAGGGACGTGGACGCAGGCGTCCAN 120
Db      61 ATCGCGAGTGCCAAAAAGTAACAATGGCTGATGATCAGGGACGTGGACGCAGGCGTCCAN 120

Qy     121 TTAACGAAGATGATTCCTCTACTTCCCGAGGTAGTGGTGATGGGCCGCGGGTGAAAGTAT 180
Db     121 TTAACGAAGATGATTCCTCTACTTCCCGAGGTAGTGGTGATGGGCCGCGGGTGAAAGTAT 180

Qy     181 TCAGAGGATCTTCATCAGGTGACCCGAGAGCGGATCCTCGTATAGAGGCTTCAAGAGAGA 240
Db     181 TCAGAGGATCTTCATCAGGTGACCCGAGAGCGGATCCTCGTATAGAGGCTTCAAGAGAGA 240

Qy     241 GAAGAGCTCTCGAGGAAGCTCCCAGGCGTGAAGGTGGCCCGCCAGAGCGAAAGCCGTGGG 300
Db     241 GAAGAGCTCTCGAGGAAGCTCCCAGGCGTGAAGGTGGCCCGCCAGAGCGAAAGCCGTGGG 300

Qy     301 GTGACCAATATGATTACCTGAATACCCGTCCGGTTGAGCTGGTATCCAAGAAGGGAACCG 360
Db     301 GTGACCAATATGATTACCTGAATACCCGTCCGGTTGAGCTGGTATCCAAGAAGGGAACCG 360

Qy     361 ATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGANTAAAAACCAAGCCGGAATGGC 420
Db     361 ATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGANTAAAAACCAAGCCGGAATGGC 420

Qy     421 GGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAGAATCCTCGTGTCCGTATGG 480
Db     421 GGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAGAATCCTCGTGTCCGTATGG 480

Qy     481 GAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTATCTATTCGACGGACTGCAAC 540
Db     481 GAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTATCTATTCGACGGACTGCAAC 540

Qy     541 TGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTCAGCGGAAAGTCGAAGCTGG 600
Db     541 TGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTCAGCGGAAAGTCGAAGCTGG 600

Qy     601 ACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATATCGTGTGCTGAGCCCCGCT 660
Db     601 ACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATATCGTGTGCTGAGCCCCGCT 660

Qy     661 TTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCGATGAAGGGCCTAAATTTGGAATTAG 720
Db     661 TTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCGATGAAGGGCCTAAATTTGGAATTAG 720

Qy     721 TTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATAAGGGAGTTCAAAATGGAGC 780
Db     721 TTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATAAGGGAGTTCAAAATGGAGC 780

Qy     781 TATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAAGATATTTTATTGGGCACCG 840
Db     781 TATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAAGATATTTTATTGGGCACCG 840

Qy     841 AAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGACATAATGCGACGTTGCTCAC 900
Db     841 AAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGACATAATGCGACGTTGCTCAC 900

Qy     901 ACAATCCGGCTCGTCATCAGGACGAAGTACGGGTAAATGTTTTGGACTTGATTGTCCTTA 960
Db     901 ACAATCCGGCTCGTCATCAGGACGAAGTACGGGTAAATGTTTTGGACTTGATTGTCCTTA 960

Qy     961 CGGATTACAATAACAGAACTTATCGTATCAATGATGTCGACTTTGGACAAACTCCGAAAT 1020
Db     961 CGGATTACAATAACAGAACTTATCGTATCAATGATGTCGACTTTGGACAAACTCCGAAAT 1020

Qy    1021 CAACATTCAGTTGCAAGGGTAGAGATATCAGTTTCGTGGAATACTATCTCACTAAATATA 1080
Db    1021 CAACATTCAGTTGCAAGGGTAGAGATATCAGTTTCGTGGAATACTATCTCACTAAATATA 1080

Qy    1081 ATATACGCATTCGCGACCACAATCAGCCGCTGCTGATTTCCAAAAATAGGGACAAGGCTC 1140
Db    1081 ATATACGCATTCGCGACCACAATCAGCCGCTGCTGATTTCCAAAAATAGGGACAAGGCTC 1140

Qy    1141 TAAAAACTAACGCTAGCGAATTAGTGGTACTAATTCCTGAGCTCTGCCGAGTGACTGGGC 1200
Db    1141 TAAAAACTAACGCTAGCGAATTAGTGGTACTAATTCCTGAGCTCTGCCGAGTGACTGGGC 1200

Qy    1201 TCAATGCCGAGATGCGCTCAAACTTTCAGCTTATGCGTGCCATGAGCAGTTATACGCGAA 1260
Db    1201 TCAATGCCGAGATGCGCTCAAACTTTCAGCTTATGCGTGCCATGAGCAGTTATACGCGAA 1260

Qy    1261 TGAACCCCAAACAACGCACTGATCGATTGCGCGCTTTTAACCACCGTTTACAAAACACTC 1320
Db    1261 TGAACCCCAAACAACGCACTGATCGATTGCGCGCTTTTAACCACCGTTTACAAAACACTC 1320

Qy    1321 CAGAAAGTGTGAAGGTCTTGAGAGACTGGAACATGGAACTGGACAAGAACGTCACAGAAG 1380
Db    1321 CAGAAAGTGTGAAGGTCTTGAGAGACTGGAACATGGAACTGGACAAGAACGTCACAGAAG 1380

Qy    1381 TACAAGGCCGGATAATTGGACAGCAGAACATCGTGTTTCATAATGGAAAGGTTCCTGCTG 1440
Db    1381 TACAAGGCCGGATAATTGGACAGCAGAACATCGTGTTTCATAATGGAAAGGTTCCTGCTG 1440

Qy    1441 GAGAAAACGCTGATTGGCAAAGGCACTTCAGAGACCAAAGGATGCTTACCACTCCGAGCG 1500
Db    1441 GAGAAAACGCTGATTGGCAAAGGCACTTCAGAGACCAAAGGATGCTTACCACTCCGAGCG 1500

Qy    1501 ATGGCCTCGATCGTTGGGCTGTCATCGCGCCGCAAAGGAATTCCCATGAACTCCGAACTC 1560
Db    1501 ATGGCCTCGATCGTTGGGCTGTCATCGCGCCGCAAAGGAATTCCCATGAACTCCGAACTC 1560

Qy    1561 TACTTGACTCTTTGTATAGAGCAGCTAGTGGAATGGGTCTTAGAATTCGAAGCCCCCAGG 1620
Db    1561 TACTTGACTCTTTGTATAGAGCAGCTAGTGGAATGGGTCTTAGAATTCGAAGCCCCCAGG 1620

Qy    1621 AATTCATAATTTATGATGATCGCACTGGAACTTATGTGAGAGCAATGGATGATTGTGTGC 1680
Db    1621 AATTCATAATTTATGATGATCGCACTGGAACTTATGTGAGAGCAATGGATGATTGTGTGC 1680

Qy    1681 GCTCAGATCCCAAACTTATATTATGCCTCGTACCCAATGATAACGCCGAAAGATACTCAT 1740
Db    1681 GCTCAGATCCCAAACTTATATTATGCCTCGTACCCAATGATAACGCCGAAAGATACTCAT 1740

Qy    1741 CAATCAAAAAGAGAGGATACGTTGACAGGGCGGTGCCAACTCAAGTTGTGACCCTTAAAA 1800
Db    1741 CAATCAAAAAGAGAGGATACGTTGACAGGGCGGTGCCAACTCAAGTTGTGACCCTTAAAA 1800

Qy    1801 CGACCAAGAACCGTAGCCTTATGAGCATTGCCACCAAAATAGCAATCCAACTGAATTGCA 1860
Db    1801 CGACCAAGAACCGTAGCCTTATGAGCATTGCCACCAAAATAGCAATCCAACTGAATTGCA 1860

Qy    1861 AGTTGGGATATACACCCTGGATGATCGAACTACCCTTGTCCGGACTGATGACAATTGGCT 1920
Db    1861 AGTTGGGATATACACCCTGGATGATCGAACTACCCTTGTCCGGACTGATGACAATTGGCT 1920

Qy    1921 TTGACATTGCGAAGAGCACACGAGATCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAA 1980
Db    1921 TTGACATTGCGAAGAGCACACGAGATCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAA 1980

Qy    1981 TGGATCTACAGCAAAACTCCACGTACTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATG 2040
Db    1981 TGGATCTACAGCAAAACTCCACGTACTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATG 2040

Qy    2041 TGCTCGCTAACACCCTTTGGCCGATGATAGCAAAGGCCCTGCGCCAATATCAACATGAGC 2100
Db    2041 TGCTCGCTAACACCCTTTGGCCGATGATAGCAAAGGCCCTGCGCCAATATCAACATGAGC 2100

Qy    2101 ATAGGAAGCTGCCATCTCGAATCGTATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAA 2160
Db    2101 ATAGGAAGCTGCCATCTCGAATCGTATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAA 2160

Qy    2161 AGCAGCTTTTTGAATTTGAAGTCAAGGACATCATTGAGAAGTTGAAAACTGAATACGCCC 2220
```

```
Db       2161 AGCAGCTTTTTGAATTTGAAGTCAAGGACATCATTGAGAAGTTGAAAACTGAATACGCCC 2220

Qy       2221 GCGTCCAGCTAAGCCCACCGCAATTAGCTTATATTGTGGTAACCAGATCCATGAACACGC 2280

Db       2221 GCGTCCAGCTAAGCCCACCGCAATTAGCTTATATTGTGGTAACCAGATCCATGAACACGC 2280

Qy       2281 GCTTCTTCCTCAACGGACAAAATCCTCCGCCTGGTACTATAGTTGATGACGTTATAACTC 2340

Db       2281 GCTTCTTCCTCAACGGACAAAATCCTCCGCCTGGTACTATAGTTGATGACGTTATAACTC 2340

Qy       2341 TGCCCGAGAGATACGACTTTTATCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGC 2400

Db       2341 TGCCCGAGAGATACGACTTTTATCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGC 2400

Qy       2401 CGACCAGCTACAATGTTCTTTATAGCAGCATGGGTNTCTCACCGGAGAAAATGCAAAAAC 2460

Db       2401 CGACCAGCTACAATGTTCTTTATAGCAGCATGGGTNTCTCACCGGAGAAAATGCAAAAAC 2460

Qy       2461 TTACGTACAAGATGTGCCACTTGTACTACAATTGGTCGGGCACCACACGAGTGCCAGCAG 2520

Db       2461 TTACGTACAAGATGTGCCACTTGTACTACAATTGGTCGGGCACCACACGAGTGCCAGCAG 2520

Qy       2521 TTTGCCAGTACGCTAAGAAGCTAGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGC 2580

Db       2521 TTTGCCAGTACGCTAAGAAGCTAGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGC 2580

Qy       2581 AAAACGCGCTCGAAAAGAAGTTTTATTATCTATAATTGGATATAATTTAGAATGGAGTAT 2640

Db       2581 AAAACGCGCTCGAAAAGAAGTTTTATTATCTATAATTGGATATAATTTAGAATGGAGTAT 2640

Qy       2641 TAATCCTTACTAAGAGGCCATATATGAAACTAGCCCAGACATTTATACTTTTTCAATACT 2700

Db       2641 TAATCCTTACTAAGAGGCCATATATGAAACTAGCCCAGACATTTATACTTTTTCAATACT 2700

Qy       2701 TCCTTACTTTTGCTAAGCACTTCAGCATTTATGACTAAATATTTTGTATTTGAAATGCAT 2760

Db       2701 TCCTTACTTTTGCTAAGCACTTCAGCATTTATGACTAAATATTTTGTATTTGAAATGCAT 2760

Qy       2761 TACTGCTCTTTTTTCAAACAAAAGCAAAATTGAGGATTAAGATTCTGGTATTTAAGCATA 2820

Db       2761 TACTGCTCTTTTTTCAAACAAAAGCAAAATTGAGGATTAAGATTCTGGTATTTAAGCATA 2820

Qy       2821 AGACCAGAGGAAATTCCCAAACAAACATTTAAAGTTATCTATCAAGACATGTTCATTAAT 2880

Db       2821 AGACCAGAGGAAATTCCCAAACAAACATTTAAAGTTATCTATCAAGACATGTTCATTAAT 2880

Qy       2881 TTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTTTATGTAAAGAAAAATTACA 2940

Db       2881 TTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTTTATGTAAAGAAAAATTACA 2940

Qy       2941 TACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGCTTATATTCGTGAAAAAAAA 3000

Db       2941 TACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGCTTATATTCGTGAAAAAAAA 3000

Qy       3001 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3047

Db       3001 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3047
```

RESULT 2
US-09-873-737A-3
; Sequence 3, Application US/09873737A
; Patent No. US20020076797A1
; GENERAL INFORMATION:
; APPLICANT: Duke University
; APPLICANT: Lin, Haifan
; TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
; TITLE OF INVENTION: PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
; FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2
; CURRENT APPLICATION NUMBER: US/09/873,737A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/US99/28764
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/110,901
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 4064
; TYPE: DNA
; ORGANISM: Mus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (191)..(2776)
; NAME/KEY: misc_feature
; LOCATION: (458)
; OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (836)
; OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (1337)
; OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (2636)
; OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
US-09-873-737A-3

Query Match 10.2%; Score 311.4; DB 10; Length 4064;
Best Local Similarity 48.7%; Pred. No. 2.3e-57;
Matches 1115; Conservative 0; Mismatches 1146; Indels 29; Gaps 9;

```
Qy        345 TCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGANTAAAA 404

Db        500 TCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCACTTCCGGCTGACC 559

Qy        405 ACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAGAAT 464

Db        560 TCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCCCTGATGGAGGCC 619

Qy        465 CCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTATCTA 524

Db        620 CGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGAAGGTGTCATGCT 679

Qy        525 TTCGACGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTCAGC 584

Db        680 TTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGTCACAGAAGTATTC 739

Qy        585 GGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATATCG 644

Db        740 AGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCACCCTGACCAACGAGCTGCCG 799

Qy        645 TGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCGATGAAGGGC 704

Db        800 CCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTTCAGGAGGCTCTTGAAAATC 859

Qy        705 CTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATAAGG 764

Db        860 ATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAGTGACCCGATTGATATTCCA 919

Qy        765 GAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAAGAT 824

Db        920 AACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCATCCTTCAGTATGAGAACAAC 979

Qy        825 ATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGACATA 884

Db        980 ATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAGCGAGACTGTCCTAGACTTC 1039

Qy        885 ATGCGACGTTGCTCACACAATCCGGCTCGTCA------TCAGGACGAAGTACGGGTAAAT 938

Db       1040 ATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCAGGAGCAAGTGTCGAAGGAG 1099

Qy        939 GTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGATGTC 998

Db       1100 CTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGACCTACCGGGTGGATGACATT 1159
```

```
Qy        999 GACTTTGGACAAACTCCGAAATCAACATTCAGTTGCAAGGGTAG---AGATATCAGTTTC 1055
Db       1160 GACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGATGGCTCGGAGGTCAGCTTC 1219

Qy       1056 GTGGAATACTATCTCACTAAATATAATATACGCATTCGCGACCACAATCAGCCGCTGCTG 1115
Db       1220 CTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGACCTGAAGCAGCCGGTGCTG 1279

Qy       1116 ATTTCCAAA---AATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTACTA 1172
Db       1280 GTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCTGCCTGGCCCAGCTATGNTC 1339

Qy       1173 ATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCAGCTT 1232
Db       1340 ATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAATGCGCAATGATTTCAATGTG 1399

Qy       1233 ATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATCGATTGCGC 1292
Db       1400 ATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCAGCGGCAGCGGGAGGTGGGC 1459

Qy       1293 GCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGACTGGAAC 1352
Db       1460 CGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAGAGAGCTTCGAGACTGGGGC 1519

Qy       1353 ATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCAGAACATC 1412
Db       1520 CTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAATCTTACAATCTGAGAAGATC 1579

Qy       1413 GTGTTTCATAATGGAAAGGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTCAGA 1472
Db       1580 ----CACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGCAGACTGGTCCAAAGA 1635

Qy       1473 GACCAAAGGATGCTTACCACTCCGAGCGATGGC-CTCGATCGTTGGGCTGTCATCGCGCC 1531
Db       1636 AACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGGCTGCTGATCTATAC 1695

Qy       1532 GCAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAGTGG 1591
Db       1696 CCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTTCAAAGTGACTCCAGC 1755

Qy       1592 AATGGGTCTTAGAATTCGAAGCCCCCAGGAATTCATAATTTATGATGATCGCACTGGAAC 1651
Db       1756 CATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGT---GGATGACAGAACAGAAGC 1812

Qy       1652 TTATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGCCTCGT 1711
Db       1813 TTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGTCTCTT 1872

Qy       1712 ACCCAATGATAACGCCGAAAGATACTCATCAATCAAAAAGAGAGGATACGTTGACAGGGC 1771
Db       1873 GTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTGTACAGACTGCCC 1932

Qy       1772 GGTGCCAACTCA---AGTTGTGACCCTTAAAACGACCAAGAACCGTAGCCTTATGAGCAT 1828
Db       1933 CACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATGGCCAT 1992

Qy       1829 TGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGATCGA 1888
Db       1993 TGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGGGTGGA 2052

Qy       1889 ACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGATCG 1948
Db       2053 CATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACACCACAGCTGG 2112

Qy       1949 GAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTT 2008
Db       2113 GCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTC 2172

Qy       2009 CAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATGAT 2068
Db       2173 CCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGCTTGCA 2232

Qy       2069 AGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATT 2128
Db       2233 AGC---TGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATCGT 2289

Qy       2129 TTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAAGGA 2188
Db       2290 GTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCACA 2349

Qy       2189 CATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATTAGC 2248
Db       2350 GTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATCGT 2409

Qy       2249 TTATATTGTGGTAACCAGATCCATGAA---CACGCGCTTCTTCCTCAACGGACAAAATCC 2305
Db       2410 GGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAACCC 2469

Qy       2306 TCCGCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTATCT 2365
Db       2470 TCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTCAT 2529

Qy       2366 GGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAG 2425
Db       2530 CGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTATGA 2589

Qy       2426 CAGCATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTA 2485
Db       2590 CAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTGTA 2649

Qy       2486 CTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGC 2545
Db       2650 CTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTGGC 2709

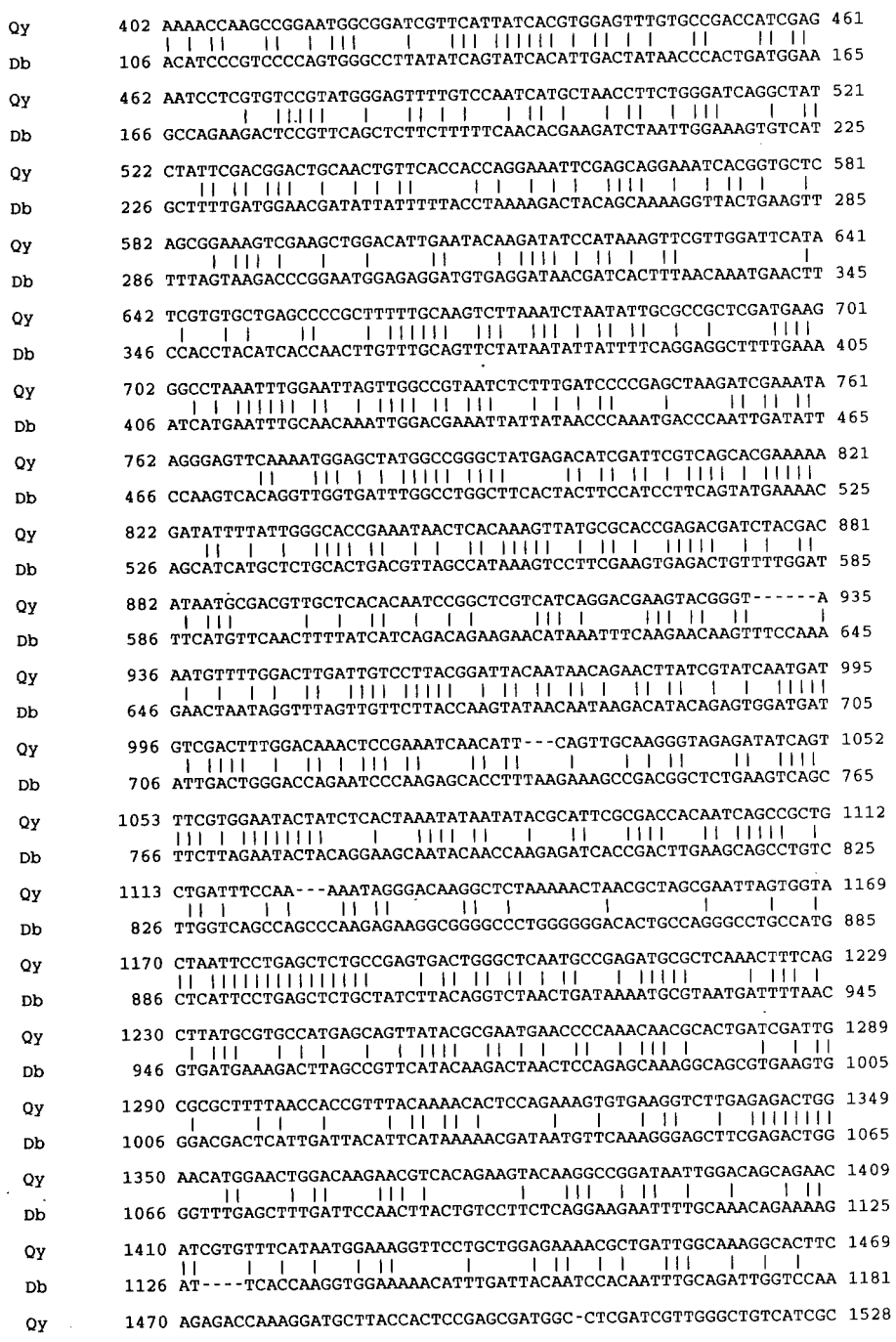
Qy       2546 TACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTA 2605
Db       2710 CTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTCTA 2769

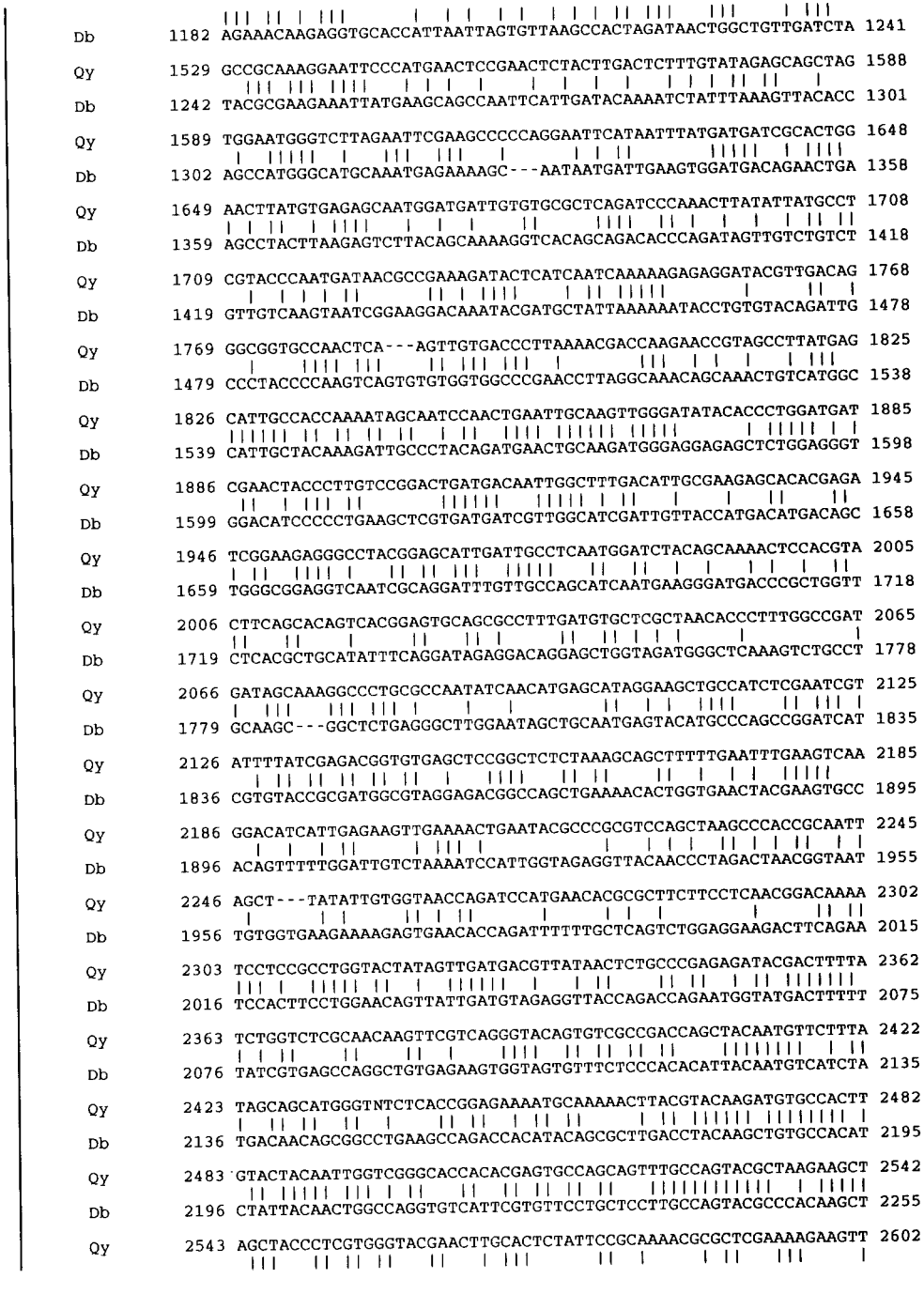
Qy       2606 TTATCTATAA 2615
Db       2770 CTACCTCTAA 2779
```

RESULT 3
US-10-043-774B-1
; Sequence 1, Application US/10043774B
; Publication No. US20030129599A1
; GENERAL INFORMATION:
; APPLICANT: University of Illinois at Chicago
; APPLICANT: Sharma, Arun
; APPLICANT: Hoffman, Ronald
; TITLE OF INVENTION: HUMAN HEMATOPOIETIC GROWTH REGULATORY GENE AND USES
; FILE REFERENCE: MBHB: CU08/PPA
; CURRENT APPLICATION NUMBER: US/10/043,774B
; CURRENT FILING DATE: 2002-06-06
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 2328
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2328)
; OTHER INFORMATION: Human Hiwi Protein
US-10-043-774B-1

Query Match 8.4%; Score 256.6; DB 14; Length 2328;
Best Local Similarity 47.2%; Pred. No. 1.4e-45;
Matches 1083; Conservative 0; Mismatches 1181; Indels 29; Gaps 9;

```
Qy        342 GTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGANTA 401
Db         46 GAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTG 105
```





```
Db       2256 GGCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCT 2315

Qy       2603 TTATTATCTATAA 2615
              ||| || || |||
Db       2316 TTACTACCTCTAA 2328


RESULT 4
US-09-873-737A-5
; Sequence 5, Application US/09873737A
; Patent No. US20020076797A1
; GENERAL INFORMATION:
;  APPLICANT: Duke University
;  APPLICANT:  Lin, Haifan
;  TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
;  TITLE OF INVENTION:  PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
;  FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2
;  CURRENT APPLICATION NUMBER: US/09/873,737A
;  CURRENT FILING DATE:  2001-06-04
;  PRIOR APPLICATION NUMBER: PCT/US99/28764
;  PRIOR FILING DATE: 1999-12-03
;  PRIOR APPLICATION NUMBER: 60/110,901
;  PRIOR FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 3472
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (167)..(2749)
;   NAME/KEY: misc_feature
;   LOCATION: (392)
;   OTHER INFORMATION: n=t or a, Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (1073)
;   OTHER INFORMATION: n=c or a, Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (2369)
;   OTHER INFORMATION: n=c or a, Xaa=Leu or Ile
US-09-873-737A-5


  Query Match             8.3%;  Score 253;  DB 10;  Length 3472;
  Best Local Similarity   47.1%;  Pred. No. 1.1e-44;
  Matches 1080;  Conservative    0;  Mismatches 1184;  Indels   29;  Gaps    9;

Qy        342 GTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGANTA 401
              | ||| || |  ||  |    ||| |    || | | |    || |||  ||| ||  |
Db        470 GAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTG 529

Qy        402 AAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAG 461
              | | ||   ||  | |||    |   ||| |||||| | || | |   ||    || ||
Db        530 ACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGGAA 589

Qy        462 AATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTAT 521
                  |  |||||   |   || | |   | || |   | ||  | |||    |  ||
Db        590 GCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCAT 649

Qy        522 CTATTCGACGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTC 581
              || || |||  |  | | ||     | |  | | |  ||||  |  | | |   |
Db        650 GCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTT 709

Qy        582 AGCGGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATA 641
              | ||| |   |   |     ||    | |||| | || |  ||           |
Db        710 TTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATGAACTT 769

Qy        642 TCGTGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCGATGAAG 701
              |   | |    ||    | ||||||  |||  ||| | || ||  |  |     ||||
Db        770 CCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAA 829

Qy        702 GGCCTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATA 761
              | | |||||| ||  | |||| || |||   | | | ||    |     || || ||

Db        830 ATCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCAATTGATATT 889

Qy        762 AGGGAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAA 821
                    ||   ||| | | ||||| ||||      || || || | |||| | |||||
Db        890 CCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAAC 949

Qy        822 GATATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGAC 881
              ||  |  |  |||| ||  | |  || |||||  | || |  |||||  | |  ||
Db        950 AGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGAT 1009

Qy        882 ATAATGCGACGTTGCTCACACAATCCGGCTCGTCATCAGGACGAAGTACGGGT------A 935
              | |||       |  |  ||  |  | |      ||| |     ||| ||  ||      |
Db       1010 TTCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAA 1069

Qy        936 AATGTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGAT 995
              |  |  | |  ||  |||| |||||  | || || |  |  || ||  |  |  |||||
Db       1070 GAANTAATAGGTTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGAGTGGATGAT 1129

Qy        996 GTCGACTTTGGACAAACTCCGAAATCAACATT---CAGTTGCAAGGGTAGAGATATCAGT 1052
              | ||||  |  || | ||| ||    || ||     |  | | ||    |  ||||
Db       1130 ATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGGGGTCAGC 1189

Qy       1053 TTCGTGGAATACTATCTCACTAAATATAATATACGCATTCGCGACCACAATCAGCCGCTG 1112
              ||| | ||||||||     |   |||| ||   |  ||   ||||   || |||||  |
Db       1190 TTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTC 1249

Qy       1113 CTGATTTCCAA---AAATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTA 1169
              || |   | |    || ||      ||  |         |       |   |   |
Db       1250 TTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATG 1309

Qy       1170 CTAATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCAG 1229
              || ||||||||||||||||     | || || || | ||   | |||||     | ||| |
Db       1310 CTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAAC 1369

Qy       1230 CTTATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATCGATTG 1289
              | |||   | | |   | | ||||  || | |  ||  | ||| |      |  | ||
Db       1370 GTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTG 1429

Qy       1290 CGCGCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGACTGG 1349
              |    | |  |     | || || |     |   |  | ||   | ||||||||
Db       1430 GGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGG 1489

Qy       1350 AACATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCAGAAC 1409
                 ||   | ||    ||| |          |  ||| | ||  |    | ||
Db       1490 GGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAG 1549

Qy       1410 ATCGTGTTTCATAATGGAAAGGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTC 1469
              ||   |  | |  | ||      |    | ||  | |  |||  |  | |
Db       1550 AT----TCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCCAA 1605

Qy       1470 AGAGACCAAAGGATGCTTACCACTCCGAGCGATGGC-CTCGATCGTTGGGCTGTCATCGC 1528
              ||| || | |||    |  | | | | | | || |||   |||     | |||
Db       1606 AGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTA 1665

Qy       1529 GCCGCAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAG 1588
              |||  |||  ||||   | | |  |  |  | |  |  | | | | ||  ||   |
Db       1666 TACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACC 1725

Qy       1589 TGGAATGGGTCTTAGAATTCGAAGCCCCCAGGAATTCATAATTTATGATGATCGCACTGG 1648
              |  |||||  |   |||  |||   |      | | ||       |||||  | ||||
Db       1726 AGCCATGGGCATGCAAATGAGAAAAGC---AATAATGATTGAAGTGGATGACAGAACTGA 1782

Qy       1649 AACTTATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGCCT 1708
              | | ||  | ||||   |  |  |    ||    |||| || |  |  |  | || ||
Db       1783 AGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCT 1842

Qy       1709 CGTACCCAATGATAACGCCGAAAGATACTCATCAATCAAAAAGAGAGGATACGTTGACAG 1768
              |  | | | ||    || | ||||   | || |||||      |    || |
Db       1843 GTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTG 1902

Qy       1769 GGCGGTGCCAACTCA---AGTTGTGACCCTTAAAACGACCAAGAACCGTAGCCTTATGAG 1825
                 |  |||| |||    || ||| |||  |       |||  | |  |  | |||
Db       1903 CCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGC 1962
```

```
Qy       1826 CATTGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGAT 1885
Db       1963 CATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGT 2022

Qy       1886 CGAACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGA 1945
Db       2023 GGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGC 2082

Qy       1946 TCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTA 2005
Db       2083 TGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTT 2142

Qy       2006 CTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGAT 2065
Db       2143 CTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCT 2202

Qy       2066 GATAGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGT 2125
Db       2203 GCAAGC---GGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCAT 2259

Qy       2126 ATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAA 2185
Db       2260 CGTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCC 2319

Qy       2186 GGACATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATT 2245
Db       2320 ACAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAAT 2379

Qy       2246 AGCT---TATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAA 2302
Db       2380 TGTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAA 2439

Qy       2303 TCCTCCGCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTA 2362
Db       2440 TCCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTT 2499

Qy       2363 TCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTA 2422
Db       2500 TATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTA 2559

Qy       2423 TAGCAGCATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTT 2482
Db       2560 TGACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACAT 2619

Qy       2483 GTACTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCT 2542
Db       2620 CTATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCT 2679

Qy       2543 AGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTT 2602
Db       2680 GGCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCT 2739

Qy       2603 TTATTATCTATAA 2615
Db       2740 TTACTACCTCTAA 2752


RESULT 5
US-09-801-574-15
; Sequence 15, Application US/09801574
; Patent No. US20020081592A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Peijing Jeremy
;  APPLICANT:  Page, David C.
;  TITLE OF INVENTION: Reproduction-Specific Genes
;  FILE REFERENCE: 0399.2007-002
;  CURRENT APPLICATION NUMBER: US/09/801,574
;  CURRENT FILING DATE:  2001-03-07
;  PRIOR APPLICATION NUMBER: 60/187,518
;  PRIOR FILING DATE: 2000-03-07
;  PRIOR APPLICATION NUMBER: 60/261,557
;  PRIOR FILING DATE: 2001-01-12
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
;   LENGTH: 3649
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-801-574-15

  Query Match             5.4%;  Score 164;  DB 10;  Length 3649;
  Best Local Similarity   45.8%;  Pred. No. 2.5e-25;
  Matches  849;  Conservative    0;  Mismatches  976;  Indels   30;  Gaps    7;

Qy        786 CCGGGCTATGAGACATCGATTCGTCAGCACGAAAAAGATATTTTATTGGGCACCGAAATA 845
Db          1 CCTGGCTATGCGGCTAGTATCCGGAGGACAGACGGGGGTCTCTTCCTGCTCGCTGATGTC 60

Qy        846 ACTCACAAAGTTATGCGCACCGAGACGATCTACGACATAATGCGACGT---TGCTCACAC 902
Db         61 TCTCATAAGGTCATTCGGAACGACTCTGTGCTGGATGTCATGCATGCTATCTACCAGCAG 120

Qy        903 AATCCGGCTCGTCATCAGGACGAAGTACGGGTAAATGTTTTGGACTTGATTGTCCTTACG 962
Db        121 AACAAGGAGCACTTCCAGGACGAGTGCAGCAAGCTTCTGGTTGGCAGCATTGTCATCACG 180

Qy        963 GATTACAATAACAGAACTTATCGTATCAATGATGTCGACTTTGGACAAACTCCGAAATCA 1022
Db        181 CGCTACAACAATCGTACCTACCGAATCGATGATGTGGACTGGAACAAGACCCCTAAAGAC 240

Qy       1023 ACAT---TCAGTTGCAAGGGTAGAGATATCAGTTTCGTGGAATACTATCTCACTAAATAT 1079
Db        241 AGCTTTGTCATGTCGGACGGGAAGGAAATCACATTCCTGGAATACTACAGCAAAAACTAT 300

Qy       1080 AATATACGCATTCGCGACCACAATCAGCCGCTGCTGAT------TTCCAAAAATAGGGAC 1133
Db        301 GGGATCACAGTCAAGGAAGATGACCAGCCGCTGCTGATCCACCGGCCCAGTGAGAGACAG 360

Qy       1134 AAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTACTAATTCCTGAGCTCTGCCGAGTG 1193
Db        361 AATAACCATGGCATGTTGCTGAAGGGCGAGATCCTGCTGCTGCCCGAGCTCTCCTTCATG 420

Qy       1194 ACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCAGCTTATGCGTGCCATGAGCAGTTAT 1253
Db        421 ACGGGGATCCCTGAGAAGATGAAGAAGGACTTCAGGGCCATGAAGGACTTGACTCAGCAG 480

Qy       1254 ACGCGAATGAACCCCAAACAACGCACTGATCGATTGCGCGCTTTTAACCACCGTTTACAA 1313
Db        481 ATTAACCTGAGCCCCAAGCAGCACCACGGTGCTTTGGAATGCCTGCTGCAGAGAATTTCA 540

Qy       1314 AACACTCCAGAAAGTGTGAAGGTCTTGAGAGACTGGAACATGGAACTGGACAAGAACGTC 1373
Db        541 CAAAACGAGACAGCCAGCAATGAGCTGACCCGCTGGGGGCTCAGTCTGCATAAAGATGTC 600

Qy       1374 ACAGAAGTACAAGGCCGGATAATTGGACAGCAGAACATCGTGTTTCATAATGGAAAGGTT 1433
Db        601 CACAAGATTGAAGGTCGGCTTCTGCCAATGGAGAGGATCAACTTAAGGAACACTTCATTT 660

Qy       1434 CCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTCAGAGACCAAAGGATGCTTACCACT 1493
Db        661 GTCACATCGGAGGGCCTGAACTGGGTTAAGGAAGTGACC------AGAGATGCTTCCATT 714

Qy       1494 CCGAGCGATGGCCTCGATCGTTGGGCTGTCATCGCGCCGCAAAGGAATTCCCATGAACTC 1553
Db        715 CTAACTATTCCCATGCATTTCTGGGCACTCTTTTATCCAAAGAGAGCAATGGACCAAGCC 774

Qy       1554 CGAACTCTACTTGACTCTTTGTATAGAGCAGCTAGTGGAATGGGTCTTAGAATTCGAAGC 1613
Db        775 AGAGAACTGGTTAACATGTTGGAAAAGATTGCCGGGCCCATTGGCATGCGCACAAGCCCC 834

Qy       1614 CCCCAGGAATTCATAATTTATGATGATCGCACTGGAACTTATGTGAGAGCAATGGATGAT 1673
Db        835 CCAGCCTGGGTTGAGCTGAAGGATGACCGAATAGAGACCTATATCAGGACCATTCAGTCC 894

Qy       1674 TGTGTGCGCTCAGATCCCA------AACTTATATTATGCCTCGTACCCAATGATAACGCC 1727
```

```
Db        895 TTACTGGGAGTTGAGGGGAAGATACAAATGGTCGTTTGCATCATCATGGGCACACGTGAT 954

Qy       1728 GAAAGATACTCATCAATCAAAAAGAGAGGATACGTTGACAGGGCGGTGCCAACTCAAGTT 1787

Db        955 GATCTCTATGGAGCCATCAAGAAGCTGTGCTGCGTGCAGTCCCCAGTGCCCTCACAGGTC 1014

Qy       1788 GTGACCCTTAAAACGA---CCAAGAACCGTAGCCTTATGAGCATTGCCACCAAAATAGCA 1844

Db       1015 ATCAATGTCCGAACCATTGGTCAGCCCACCAGGCTTCGGAGCGTGGCTCAGAAAATTTTA 1074

Qy       1845 ATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGATCGAACTACCCTTGTCCGGA 1904

Db       1075 CTTCAGATGAACTGTAAACTGGGTGGTGAGCTCTGGGGAGTGGATATTCCGCTGAAACAA 1134

Qy       1905 CTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGATCGGAAGAGGGCCTACGGA 1964

Db       1135 CTAATGGTGATTGGAATGGATGTGTACCATGACCCCAGCAGAGGCATGCGCTCTGTGGTC 1194

Qy       1965 GCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTTCAGCACAGTCACGGAG 2024

Db       1195 GGCTTCGTGGCCAGCATAAATCTCACACTCACCAAATGGTACTCGAGGGTGGTGTTCCAG 1254

Qy       2025 TGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATGATAGCAAAGGCCCTGCGC 2084

Db       1255 ---ATGCCACATCAGGAGATTGTGGACAGCCTGAAGCTCTGCCTGGTGGGTTCCTTGAAA 1311

Qy       2085 CAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATTTTATCGAGACGGTGTG 2144

Db       1312 AAGTATTATGAGGTGAACCATTGTCTCCCAGAGAAAATTGTGGTGTACCGAGATGGAGTG 1371

Qy       2145 AGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAAGGACATCATTGAGAAGTTG 2204

Db       1372 TCTGATGGCCAGCTAAAGACAGTTGCCAACTACGAGATCCCTCAGCTGCAGAAGTGTTTT 1431

Qy       2205 AAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATTAGCTTATATTGTGGTAACC 2264

Db       1432 GAAGCCTTTGATAACTACCACCCCAAGATGGTGGTGTTTGTAGTTCAGAAGAAAATCAGC 1491

Qy       2265 AGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAATCCTCCGCCTGGTACTATAGTT 2324

Db       1492 ACCAATCTGTACCTTGCTGCTCCTGATCACTTCGTAACCCCCTCCCCCGGGACTGTGGTT 1551

Qy       2325 GATGACGTTATAACTCTGCCCGAGAGATACGACTTTTATCTGGTCTCGCAACAAGTTCGT 2384

Db       1552 GATCATACCATAACCAGCTGTGAGTGGGTGGATTTCTACCTTCTTGCCCATCATGTGCGA 1611

Qy       2385 CAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAGCAGCATGGGTNTCTCACCG 2444

Db       1612 CAGGGCTGTGGCATACCTACACACTACATCTGTGTTCTGAACACTGCAAATCTGAGCCCT 1671

Qy       2445 GAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTACTACAATTGGTCGGGCACC 2504

Db       1672 GATCACATGCAGAGGTTGACTTTCAAACTATGCCACATGTACTGGAATTGGCCTGGTACC 1731

Qy       2505 ACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGCTACCCTCGTGGGTACGAAC 2564

Db       1732 ATCCGAGTTCCAGCTCCTTGCAAGTATGCCCACAAGCTAGCTTTCCTGTCCGGACAGATT 1791

Qy       2565 TTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTATTATCTATAATTGG 2619

Db       1792 TTGCATCATGAGCCAGCCATCCAGCTGTGTGGGAACCTGTTCTTCCTGTAACTGG 1846


RESULT 6
US-09-998-598-2602
; Sequence 2602, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
;  APPLICANT: Stolk, John A.
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Chenault, Ruth A.
;  APPLICANT:  Meagher, Madelein Joy
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.561
;  CURRENT APPLICATION NUMBER: US/09/998,598
;  CURRENT FILING DATE:  2001-11-16
;  NUMBER OF SEQ ID NOS: 2606
;  SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 2602
;   LENGTH: 2292
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-998-598-2602

  Query Match             5.3%;  Score 161.6;  DB 11;  Length 2292;
  Best Local Similarity   47.3%;  Pred. No. 6.4e-25;
  Matches  688;  Conservative    0;  Mismatches  750;  Indels   17;  Gaps    6;

Qy       1168 TACTAATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTC 1227

Db        128 TGCTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTA 187

Qy       1228 AGCTTATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATCGAT 1287

Db        188 ACGTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAG 247

Qy       1288 TGCGCGCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGACT 1347

Db        248 TGGGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACT 307

Qy       1348 GGAACATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCAGA 1407

Db        308 GGGGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAA 367

Qy       1408 ACATCGTGTTTCATAATGGAAAGGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACT 1467

Db        368 AGAT----TCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCC 423

Qy       1468 TCAGAGACCAAAGGATGCTTACCACTCCGAGCGATGGC-CTCGATCGTTGGGCTGTCATC 1526

Db        424 AAAGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATC 483

Qy       1527 GCGCCGCAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCT 1586

Db        484 TATACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACA 543

Qy       1587 AGTGGAATGGGTCTTAGAATTCGAAGCCCCCAGGAATTCATAATTTATGATGATCGCACT 1646

Db        544 CCAGCCATGGGCATGCAAATGAGAAAAGC---AATAATGATTGAAGTGGATGACAGAACT 600

Qy       1647 GGAACTTATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGC 1706

Db        601 GAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGT 660

Qy       1707 CTCGTACCCAATGATAACGCCGAAAGATACTCATCAATCAAAAAGAGAGGATACGTTGAC 1766

Db        661 CTGTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGAT 720

Qy       1767 AGGGCGGTGCCAACTCA---AGTTGTGACCCTTAAAACGACCAAGAACCGTAGCCTTATG 1823

Db        721 TGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATG 780

Qy       1824 AGCATTGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATG 1883

Db        781 GCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGG 840

Qy       1884 ATCGAACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGA 1943

Db        841 GTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACA 900

Qy       1944 GATCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACG 2003

Db        901 GCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGG 960
```

```
Qy       2004 TACTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCG 2063
Db        961 TTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGC 1020

Qy       2064 ATGATAGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATC 2123
Db       1021 CTGCAAGC---GGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATC 1077

Qy       2124 GTATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTC 2183
Db       1078 ATCGTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTG 1137

Qy       2184 AAGGACATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAA 2243
Db       1138 CCACAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGACTAACGGTA 1197

Qy       2244 ---TTAGCTTATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACAA 2300
Db       1198 ATTGTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAG 1257

Qy       2301 AATCCTCCGCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTT 2360
Db       1258 AATCCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTT 1317

Qy       2361 TATCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTT 2420
Db       1318 TTTATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATC 1377

Qy       2421 TATAGCAGCATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCAC 2480
Db       1378 TATGACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCAC 1437

Qy       2481 TTGTACTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAG 2540
Db       1438 ATCTATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAG 1497

Qy       2541 CTAGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAG 2600
Db       1498 CTGGCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGC 1557

Qy       2601 TTTTATTATCTATAA 2615
Db       1558 CTTTACTACCTCTAA 1572
```

```
RESULT 7
US-10-029-495-8
; Sequence 8, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
;  APPLICANT: Drmanac, Radoje T.
;  APPLICANT:  Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Dickson, Mark C.
;  APPLICANT:  Jones, Lee William
;  APPLICANT:  Ballinger, Dennis G.
;  APPLICANT:  Xue, Aidong
;  APPLICANT:  Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi,Vinod
;  TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
;  FILE REFERENCE: 28110/35905A
;  CURRENT APPLICATION NUMBER: US/10/029,495
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: 09/378,667
;  PRIOR FILING DATE: 1999-08-20
;  PRIOR APPLICATION NUMBER: 09/687,527
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 2009
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (73)..(1569)
US-10-029-495-8

  Query Match             4.9%;  Score 149;  DB 15;  Length 2009;
  Best Local Similarity   46.2%;  Pred. No. 3.3e-22;
  Matches  753;  Conservative    0;  Mismatches  856;  Indels   21;  Gaps    7;

Qy       1056 GTGGAATACTATCTCACTAAATATAATATACGCATTCGCGACCACAATCAGCCGCTGCTG 1115
Db         19 GTGGATTACTACAAGCAGCAGTATGATATTACTGTATCGGACCTGAATCAGCCCATGCTT 78

Qy       1116 ATTTCCAAAAATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTACTAATT 1175
Db         79 GTTAGTCTGTTAAAGAAGAAGAGAAATGACAACAGTGAGCCTCAGCTCGCCCACCTGATA 138

Qy       1176 CCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCAGCTTATG 1235
Db        139 CCTGAGCTCTGCTTTCTAACAGGGCTGACTGACCAGGCAACATCTGATTTCCAGCTGATG 198

Qy       1236 CGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATCGATTGCGCGCT 1295
Db        199 AAGGCTGTGGCTGAAAAGACACGTCTCAGTCCTTCAGGCCGGCAGCAGCGCCTGGCCAGG 258

Qy       1296 TTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGACTGGAACATG 1355
Db        259 CTTGTGGACAACATCCAGAGGAATACCAATGCTCGCTTTGAACTAGAGACCTGGGGACTG 318

Qy       1356 GAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCAGAACATCGTG 1415
Db        319 CATTTTGGAAGCCAGATATCTC---TGACTGGCCGGATTGTGCCTTCAGAAAAAATATTA 375

Qy       1416 TTTCATAATGGAAAGGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTCAGAGAC 1475
Db        376 ATGCAAGACCACATATGTCAACCTGTGTCTGCTGCTGACTGGTCCAAGGATATTCGAACT 435

Qy       1476 CAAAGGATGCTTACCACTCCGAGCGATGGCCTCGATCGTTGGGCTGTCATCGCGCCGCAA 1535
Db        436 TGCAAGATTTTAAATGCAC--AGTCTTTGAATACCTGGTTGATTTTATGTAGCGACAGAA 493

Qy       1536 AGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAGTGGAATG 1595
Db        494 CTGAAT----ATGTTGCCGAGAGCTTTCTGAACTGCTTGAGAAGAGTTGCAGGTTCCATG 549

Qy       1596 GGTCTTAGAATTCGAAGCCCCCAGGAATTCATAATTTATGATGATCGCACTGGAACTTAT 1655
Db        550 GGATTTAATGTGGACTACCCCAAAATCATAAAAGTACAAGAAAATCCAGCTG---CATTT 606

Qy       1656 GTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGCCTCGTACCC 1715
Db        607 GTTAGAGCTATACAGCAATATGTTGATCCTGATGTTCAGCTGGTAATGTGCATTCTGCCT 666

Qy       1716 AATGATAACGCCGAAAGATACTCATCAATCAAAAAGAGAGGATACGTTGACAGGGCGGTG 1775
Db        667 TCTAATCAGAAGACCTATTATGATTCCATTAAAAAATATTTGAGCTCAGACTGCCCAGTC 726

Qy       1776 CCAACTCAAGTTGTG---ACCCTTAAAACGACCAAGAACCGTAGCCTTATGAGCATTGCC 1832
Db        727 CCAAGCCAATGTGTGCTTGCTCGGACCTTGAATAAACAGGGCATGATGATGAGTATCGCC 786

Qy       1833 ACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGATCGAACTA 1892
Db        787 ACCAAGATCGCTATGCAGATGACTTGCAAGCTCGGAGGCGAGCTGTGGGCTGTGGAAATA 846

Qy       1893 CCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGATCGGAAG 1952
Db        847 CCTTTAAAGTCCCTGATGGTGGTCGGTATTGATGTCTGTAAAGATGCACTCAGCAAGGAC 906

Qy       1953 AGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTTCAGC 2012
Db        907 GTGATGGTTGTTGGATGCGTGGCCAGTGTTAACCCCAGAATCACCAGGTGGTTTTCCCGC 966
```

```
Qy       2013 ACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATGATAGCA 2072
Db        967 TGTATCCTTCAGAGAACAATGACTGATGTTGCAGATTGCTTGAAAGTTTTCATGACTGGA 1026

Qy       2073 AAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATTTTAT 2132
Db        1027 ---GCACTCAACAAATGGTACAAGTACAATCATGATTTGCCAGCACGGATAATTGTGTAC 1083

Qy       2133 CGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAAGGACATC 2192
Db        1084 CGTGCTGGTGTAGGGGATGGTCAGCTGAAAACACTTATTGAATATGAAGTCCCACAGCTG 1143

Qy       2193 ATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATTAGCTTAT 2252
Db        1144 CTGAGCAGTGTGGCAGAATCCAGCTCAAATACCAGCTCAAGACTGTCGGTGATTGTGGTC 1203

Qy       2253 ATTGTGGTAACCAGATCCATGAAC---ACGCGCTTCTTCCTCAACGGACAAAATCCTCCG 2309
Db        1204 AGGAAGAAGTGCATGCCACGATTCTTTACCGAAATGAACCGCACTGTACAGAACCCCCCA 1263

Qy       2310 CCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTATCTGGTC 2369
Db        1264 CTTGGCACTGTTGTGGATTCAGAAGCAACACGTAACGAATGGTATGACTTTTATCTGATC 1323

Qy       2370 TCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAGCAGC 2429
Db        1324 AGCCAGGTGGCCTGCCGGGGAACTGTTAGTCCTACCTACTATAATGTCATCTATGATGAC 1383

Qy       2430 ATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTACTAC 2489
Db        1384 AACGGCTTGAAGCCCGACCATATGCAGAGACTTACATTCAAATTGTGCCACCTGTACTAC 1443

Qy       2490 AATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGCTACC 2549
Db        1444 AACTGGCCGGGCATAGTCAGTGTCCCAGCACCATGTCAGTATGCTCACAAGCTGACCTTT 1503

Qy       2550 CTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTATTAT 2609
Db        1504 CTGGGGGCACAAAGCATTCATAAAGAACCCAGTCTGGAATTAGCCAACCATCTCTTCTAC 1563

Qy       2610 CTATAATTGGATATAATTTAGAATGGAGTATTAATCCTTACTAAGAGGCCATATATGAAA 2669
Db        1564 CTGTGATGGCATGAACTACTGGCATCACTAGATGGACAATCCAAGAAGAAATTGGTATAC 1623

Qy       2670 CTAGCCCAGA 2679
Db        1624 TTTGTGCAAA 1633
```

RESULT 8
US-09-960-352-4582/c
; Sequence 4582, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 4582
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 20-LIB3057-013-Q1-K1-E7
US-09-960-352-4582

Query Match 2.7%; Score 81.2; DB 11; Length 393;
Best Local Similarity 56.9%; Pred. No. 7.6e-08;
Matches 149; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

```
Qy       2786 AAAATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAA 2845
Db        393 AAATTAAAAAAATAAAAAATTAATTTAAAAAATTAATATAAAAAAAAAAATTTATAAAAA 334

Qy       2846 CATTTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTT 2905
Db        333 TATTTAAATTTATAATAATAAAATAATATAATCATAAAATATAATAATATATTTTAAATA 274

Qy       2906 ATTGTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACC 2965
Db        273 ATTTTTTAAAAAATATATATTTAAAAAAAAAAAAAAAAATTTTATAAATAATTTTTTAAAA 214

Qy       2966 ACACATTCATGGCTGCTTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3025
Db        213 ATAAAAAAATTATAAAATAAATAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 154

Qy       3026 AAAAAAAAAAAAAAAAAAAAA 3047
Db        153 AAAAAAAAAAAAAAAAAAAAA 132
```

RESULT 9
US-09-960-352-13784/c
; Sequence 13784, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 13784
; LENGTH: 397
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 59-LIB3057-006-Q1-K1-G4
US-09-960-352-13784

Query Match 2.5%; Score 77; DB 11; Length 397;
Best Local Similarity 63.6%; Pred. No. 6.2e-07;
Matches 133; Conservative 0; Mismatches 75; Indels 1; Gaps 1;

```
Qy       2839 AAACAAACATTTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTT 2898
Db        289 ATATATTCTATTATTGTTCTCACTCAATACTTTATCATTCA-GTATATTAAAATTATTTT 231

Qy       2899 ATTTTTTATTGTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTA 2958
Db        230 TTTTATCATTATAAAAAAAATTTTTTTATAAAACAAAAAAATTTTATTAATATATGAATT 171

Qy       2959 CTTAACCACACATTCATGGCTGCTTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAA 3018
Db        170 CTTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 111

Qy       3019 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3047
Db        110 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 82
```

RESULT 10
US-09-960-352-11234/c
; Sequence 11234, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.

```
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11234
;   LENGTH: 419
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 48-LIB3058-052-Q1-K1-D8
US-09-960-352-11234

  Query Match             2.5%;  Score 76.2;  DB 11;  Length 419;
  Best Local Similarity   52.3%;  Pred. No. 9.6e-07;
  Matches  168;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy       2727 ATTTATGACTAAATATTTTGTATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGCA 2786
             | ||||   || | |  |  | ||| |||    ||  |     ||||| |   |  |
Db        371 AATTATTTTTATAAAAATATTTTTTAAAAAATTTTTTTTAAAATTTTTTATTTATTATTT 312

Qy       2787 AAATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAAC 2846
             | || |    |||  || | |  |  | ||    |   || ||   |||  |
Db        311 TATTTAAATTATAATTTTTTAAAAAATTTTAAAAATTTAATTTAAATTTAAAAAAATATA 252

Qy       2847 ATTTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTA 2906
             || ||||   || | |    | || |     ||||   ||| | |    ||||   |||
Db        251 ATATAAAAAAATTTTTTTTTATATTTAATAAAATTATAAATTTTAACTTTTTAAAATTTT 192

Qy       2907 TTGTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCA 2966
             ||   |||| |  || |   ||| |||||  | | | |   |             ||  |
Db        191 TTTATTATTATTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAA 132

Qy       2967 CACATTCATGGCTGCTTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3026
              | |   |         | |      ||||||||||||||||||||||||||||||||||
Db        131 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 72

Qy       3027 AAAAAAAAAAAAAAAAAAAA 3047
              ||||||||||||||||||||
Db         71 AAAAAAAAAAAAAAAAAAAA 51


RESULT 11
US-09-960-352-15014/c
; Sequence 15014, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 15014
;   LENGTH: 375
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 64-LIB3058-048-Q1-K1-H8
US-09-960-352-15014

  Query Match             2.5%;  Score 75.2;  DB 11;  Length 375;
  Best Local Similarity   53.0%;  Pred. No. 1.5e-06;
  Matches  161;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy       2744 TTGTATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGCAAAATTGAGGATTAAGAT 2803
             || | |||  |||    |  || | |  |||     |||| |   ||  |  |||    |

Db        366 TTTTTTTTTGAATTTTATTTTGTTATCATTTGTTTAAAAATCTTTATATAAAATTTTTTT 307

Qy       2804 TCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAACATTTAAAGTTATCTATC 2863
             | |  | |||  |  ||         |  | | |  .          ||  |||    || |   |
Db        306 TTTTATTTTTGAATATTTTTATTCATCACAGTATTTTTTTAAATTTTTTTTTTTTTATTT 247

Qy       2864 AAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTTTA 2923
                  || ||  |||  || | ||    | ||||||||  || |  | ||||  |||
Db        246 TTTTTTCCATCCTTTTTTTTAAAAAAAAAATATATATTTTTTTTTTTTTTTTTTATTTT 187

Qy       2924 TGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGCTT 2983
             |  ||| ||||| ||   | |   |   |   | |   ||  |   |   |
Db        186 TTAAAATAAAAAATAATAAAAATAAAAATAAATAAAAAAAATAATAAAAAAAAAAAAAAA 127

Qy       2984 ATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3043
             | |       |||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        126 AAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 67

Qy       3044 AAAA 3047
              ||||
Db         66 AAAA 63


RESULT 12
US-09-960-352-6976/c
; Sequence 6976, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 6976
;   LENGTH: 337
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 30-LIB3058-018-Q1-K1-H5
US-09-960-352-6976

  Query Match             2.4%;  Score 74.2;  DB 11;  Length 337;
  Best Local Similarity   56.9%;  Pred. No. 2.3e-06;
  Matches  136;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy       2809 TATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAACATTTAAAGTTATCTATCAAGAC 2868
             | |||||   ||| |        ||||   |||   | |   ||||   | | | | |
Db        305 TTTTTAAATTTAAAAAATTTTATAATTGAAAAAATTATAAACAAAGCATTTTTTTTAAAA 246

Qy       2869 ATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTTTATGTAA 2928
             || |  |  || |   | | |   |   |  |||||  || |  |   ||   | | | |
Db        245 ATTTATAAAAAATAACATTTTTCATTAATATTTTTTATTTTTTCAAAATATAATTTATTA 186

Qy       2929 AGAAAAATTACATACATCCATGTTTGCTTACTTAACCACACATTCATGGCTGCTTATATT 2988
             | ||||| ||   |    | | ||   |||   ||  |    || ||       | | |
Db        185 AAAAAAAATAAGAAATAACTTCTTATTTTAAAAAAAAAATTTTTAATTATAAAAAAAAAT 126

Qy       2989 CGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3047
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 67


RESULT 13
US-10-198-846-8560/c
; Sequence 8560, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT:  Xu, Yongyao
```

```
;  APPLICANT:  Wang, Youzhen
;  APPLICANT:  Steinmann, Kathleen
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
;  TITLE OF INVENTION:  FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF BREAST CANCER
;  FILE REFERENCE: MRI-049
;  CURRENT APPLICATION NUMBER: US/10/198,846
;  CURRENT FILING DATE:  2002-07-18
;  PRIOR APPLICATION NUMBER: 60/306,220
;  PRIOR FILING DATE: 2001-07-18
;  NUMBER OF SEQ ID NOS: 14084
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8560
;   LENGTH: 638
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 5, 7, 20, 22, 32, 37, 44, 47, 218, 219, 220, 221, 223, 257,
;   LOCATION: 258, 259, 275, 286, 287, 288, 289, 290, 293, 294, 295, 296,
;   LOCATION: 301, 302, 303, 315, 316, 317, 318, 332, 333, 334, 335, 342,
;   LOCATION: 343, 344, 359, 366, 393, 394, 395, 396, 397, 399, 400
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 402, 415, 420, 421, 422, 423, 427, 443, 444, 452, 464, 465,
;   LOCATION: 466, 467, 468, 480, 481, 482, 483, 484, 485, 486, 487, 494,
;   LOCATION: 495, 497, 498, 502, 503, 510, 511, 531, 535, 537, 542, 543,
;   LOCATION: 544, 545, 548, 549, 550, 554, 555, 556, 558, 561, 562
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 574, 577,
;   LOCATION: 578, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602,
;   LOCATION: 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614,
;   LOCATION: 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637,
;   LOCATION: 638
;   OTHER INFORMATION: n = A,T,C or G
US-10-198-846-8560

  Query Match             2.4%;  Score 73.6;  DB 15;  Length 638;
  Best Local Similarity   54.5%;  Pred. No. 4.4e-06;
  Matches  109;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy       2848 TTTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTAT 2907
             |||    ||    |  ||    | ||  ||  |||      |  ||  ||
Db        353 TTTTTTTTTNNNAAAAAANNNNTTTTTTTTTTTTTTNNNNCTTTTTTTTTTNNNAAAANNN 294

Qy       2908 TGTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCAC 2967
                ||  ||| |  ||| ||||  | |    |   | |||  ||  || | |
Db        293 NCCNNNNNTTTTTTTTTNAAAAAAAAAAAAAAANNNTTTTTTTTTTTTTTTTTTAAAAA 234

Qy       2968 ACATTCATGGCTGCTTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3027
             |  ||  |       | |      ||||||||||||||||||||||||||||||||||
Db        233 AATTTTTTTTNGNNNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 174

Qy       3028 AAAAAAAAAAAAAAAAAAA 3047
              |||||||||||||||||||
Db        173 AAAAAAAAAAAAAAAAAAA 154


RESULT 14
US-09-925-297-307
; Sequence 307, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA105
;  CURRENT APPLICATION NUMBER: US/09/925,297
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05989
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 928
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 307
;   LENGTH: 1046
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (4)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (14)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (946)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (948)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-925-297-307

  Query Match             2.4%;  Score 73.4;  DB 10;  Length 1046;
  Best Local Similarity   59.8%;  Pred. No. 6.3e-06;
  Matches  119;  Conservative    1;  Mismatches   79;  Indels    0;  Gaps    0;

Qy       2849 TTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTTATTTTTTATT 2908
             | | | ||  || |    | ||  | |    | | |  || |||     |   |||| ||
Db        805 TAATACTTTCCTGTATTTATATCATAACGTGTATAGTGTAAAATGTGAATGACTTTTTTT 864

Qy       2909 GTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTACTTAACCACA 2968
             ||  ||   | | ||          || |   || |||    |||| ||    ||  | |
Db        865 GTGAATGAAAATCTAAAATCTTTGTAACTTTTTATATCTGCTTTTGTTTCACCAAAGAAA 924

Qy       2969 CATTCATGGCTGCTTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3028
             | |  |   || ||| || :    ||||||||||||||||||||||||||||||||||||
Db        925 CCTAAAATCCTTCTTTTAMWANANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 984

Qy       3029 AAAAAAAAAAAAAAAAAA 3047
              ||||||||||||||||||
Db        985 AAAAAAAAAAAAAAAAAA 1003


RESULT 15
US-09-834-975-504/c
; Sequence 504, Application US/09834975
; Patent No. US20020110815A1
; GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT:  Brown, Jeffrey
;  APPLICANT:  Bolt, Andrew
;  APPLICANT:  Van Huffel, Christophe
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS
;  TITLE OF INVENTION:  FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
;  TITLE OF INVENTION:  OF HUMAN CANCERS
;  FILE REFERENCE: MRI-016B
;  CURRENT APPLICATION NUMBER: US/09/834,975
;  CURRENT FILING DATE:  2001-04-13
;  PRIOR APPLICATION NUMBER: 60/197,538
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 1046
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 504
;   LENGTH: 509
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-834-975-504
```

Query Match 2.4%; Score 72.8; DB 11; Length 509;
Best Local Similarity 59.0%; Pred. No. 5.8e-06;
Matches 125; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

```
Qy       2836 CCCAAACAAACATTTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTAC 2895
              || | | ||   ||||||    |  || ||   |  ||||| || | |  || |   |
Db        223 CCAACAAAAGGTTTTAAACCCTTTAATGAATTGACTTTCATAAAATGGTTATGTTGGTGG 164

Qy       2896 TTTATTTTTTATTGTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGC 2955
                | |||||   |       | ||  ||        ||||    | ||| ||  |
Db        163 GCAAGTTTTTTAAGCTGGAAATTGTAAATTCCCCCTGAAATGTTTTTCATGCAGTTCCCA 104

Qy       2956 TTACTTAACCACACATTCATGGCTGCTTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAA 3015
              | |  |||     || | | || || || |    ||   ||||||||||||||||||||
Db        103 TGAACTAATACTCCAATAAAGGATGGTTTTGGGGGTCCCAAAAAAAAAAAAAAAAAAAAA 44

Qy       3016 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3047
              ||||||||||||||||||||||||||||||
Db         43 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 12
```

Search completed: July 22, 2003, 06:59:22
Job time : 592.209 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2003, 13:46:42 ; Search time 4001.44 Seconds
(without alignments)
12332.482 Million cell updates/sec

Title: US-09-873-737A-1
Perfect score: 3047
Sequence: 1 ctgagtccaaagcgtcgttt.....aaaaaaaaaaaaaaaaaaa 3047

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	753.4	24.7	757	9	AA978826	AA978826 LD33002.3
	2	678.2	22.3	683	13	BI354756	BI354756 GM28157.5
	3	614.8	20.2	620	9	AA802575	AA802575 GM05853.5
c	4	583.8	19.2	589	12	BG637778	BG637778 LD03429.3
	5	566	18.6	567	9	AA202789	AA202789 LD03429.5
	6	559.4	18.4	573	13	BI214891	BI214891 RE21038.5
	7	521.6	17.1	538	13	BI372643	BI372643 RE59653.5
c	8	496.8	16.3	500	9	AI061964	AI061964 LD35072.3
	9	486.4	16.0	498	13	BI241150	BI241150 RE38167.5
c	10	468.8	15.4	489	12	BG637781	BG637781 LD03506.3
	11	409.4	13.4	433	13	BI168498	BI168498 RE09253.5
	12	395.6	13.0	988	17	CNS008NM	AL052102 Drosophil
	13	375.6	12.3	385	13	BI163833	BI163833 RE03347.5
c	14	325.2	10.7	410	10	AW941950	AW941950 GM05853.3
	15	196.8	6.5	775	9	AA817640	AA817640 LD24150.5
	16	154.2	5.1	642	13	BM612785	BM612785 170006871
	17	137.4	4.5	711	13	BM646472	BM646472 170006873
c	18	127.2	4.2	568	9	AI455313	AI455313 LD23107.3
	19	125.8	4.1	733	13	BM656277	BM656277 170006873
	20	125.2	4.1	683	13	BM602738	BM602738 170006870
	21	123.6	4.1	741	9	AU003278	AU003278 AU003278
	22	115.8	3.8	609	14	BQ481304	BQ481304 faa74e09.
c	23	112.8	3.7	654	9	AA695688	AA695688 GM03296.5
	24	110.8	3.6	778	10	AV401452	AV401452 AV401452
	25	107.2	3.5	829	12	BG192739	BG192739 RST11856
	26	107	3.5	472	14	BQ481269	BQ481269 faa74a08.
	27	106.2	3.5	700	14	BM716667	BM716667 UI-E-EJ0-
	28	106.2	3.5	776	12	BG195785	BG195785 RST14980
	29	106.2	3.5	783	12	BG198827	BG198827 RST18097
	30	106.2	3.5	795	12	BG191694	BG191694 RST10790
	31	106	3.5	632	12	BG193278	BG193278 RST12406
	32	106	3.5	748	12	BG211839	BG211839 RST31411
	33	106	3.5	800	12	BG206124	BG206124 RST25559
	34	105.2	3.5	796	12	BG209773	BG209773 RST29299
	35	105	3.4	768	12	BG202422	BG202422 RST21779
	36	105	3.4	785	12	BG221179	BG221179 RST40982
	37	104.8	3.4	750	12	BG191843	BG191843 RST10945
	38	104.4	3.4	777	12	BG185416	BG185416 RST4359 A
	39	104.4	3.4	778	12	BG210801	BG210801 RST30349
	40	102.6	3.4	385	9	AI986645	AI986645 rs15f02.y
	41	102.6	3.4	440	10	AW052355	AW052355 rs34e05.y
	42	102.6	3.4	598	9	AA193873	AA193873 rs06d11.r
	43	99	3.2	649	13	BM611141	BM611141 170006871
	44	98.4	3.2	635	10	AW941193	AW941193 GM03296.3
	45	97	3.2	749	12	BG215043	BG215043 RST34701

ALIGNMENTS

RESULT 1
AA978826/c
LOCUS AA978826 757 bp mRNA linear EST 19-APR-2001
DEFINITION LD33002.3prime LD Drosophila melanogaster embryo pOT2 Drosophila melanogaster cDNA clone LD33002 3prime, mRNA sequence.
ACCESSION AA978826
VERSION AA978826.1 GI:3156242
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 757)
AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G., Lewis,S. and Rubin,G.M.
TITLE BDGP/HHMI Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Based upon the presence of a XhoI site followed by a run of 14 or more T residues at the beginning of the sequence, this clone probably contains an inverted insert. The sequence has been trimmed and the T residues removed.
Plate: 330 row: A column: 2

```
          High quality sequence stop: 667.
FEATURES             Location/Qualifiers
     source          1. .757
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LD33002"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
BASE COUNT      235 a    128 c    158 g    236 t
ORIGIN

  Query Match             24.7%;  Score 753.4;  DB 9;  Length 757;
  Best Local Similarity   99.7%;  Pred. No. 1.5e-101;
  Matches  754;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2210 TGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATTAGCTTATATTGTGGTAACCAGATC 2269
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        756 TGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATTAGCTTATATTGTGGTAACCAGATC 697

Qy       2270 CATGAACACGCGCTTCTTCCTCAACGGACAAAATCCTCCGCCTGGTACTATAGTTGATGA 2329
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         696 CATGAACACGCGCTTCTTCCTCAACGGACAAAATCCTCCGCCTGGTACTATAGTTGATGA 637

Qy       2330 CGTTATAACTCTGCCCGAGAGATACGACTTTTATCTGGTCTCGCAACAAGTTCGTCAGGG 2389
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         636 CGTTATAACTCTGCCCGAGAGATACGACTTTTATCTGGTCTCGCAACAAGTTCGTCAGGG 577

Qy       2390 TACAGTGTCGCCGACCAGCTACAATGTTCTTTATAGCAGCATGGGTNTCTCACCGGAGAA 2449
             |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db         576 TACAGTGTCGCCGACCAGCTACAATGTTCTTTATAGCAGCATGGGTCTCTCACCGGAGAA 517

Qy       2450 AATGCAAAAACTTACGTACAAGATGTGCCACTTGTACTACAATTGGTCGGGCACCACACG 2509
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         516 AATGCAAAAACTTACGTACAAGATGTGCCACTTGTACTACAATTGGTCGGGCACCACACG 457

Qy       2510 AGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGCTACCCTCGTGGGTACGAACTTGCA 2569
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         456 AGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGCTACCCTCGTGGGTACGAACTTGCA 397

Qy       2570 CTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTATTATCTATAATTGGATATAATTTA 2629
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         396 CTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTATTATCTATAATTGGATATAATTTA 337

Qy       2630 GAATGGAGTATTAATCCTTACTAAGAGGCCATATATGAAACTAGCCCAGACATTTATACT 2689
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         336 GAATGGAGTATTAATCCTTACTAAGAGGCCATATATGAAACTAGCCCAGACATTTATACT 277

Qy       2690 TTTTCAATACTTCCTTACTTTTGCTAAGCACTTCAGCATTTATGACTAAATATTTTGTAT 2749
             |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db         276 TTTTCAATACTTTCTTACTTTTGCTAAGCACTTCAGCATTTATGACTAAATATTTTGTAT 217

Qy       2750 TTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGCAAAATTGAGGATTAAGATTCTGGT 2809
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         216 TTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGCAAAATTGAGGATTAAGATTCTGGT 157

Qy       2810 ATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAACATTTAAAGTTATCTATCAAGACA 2869
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         156 ATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAACATTTAAAGTTATCTATCAAGACA 97

Qy       2870 TGTTCATTAATTTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTTTATGTAAA 2929
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          96 TGTTCATTAATTTGGAATATAATTACTTTATTTTTTATTGTATATTTTAGTTTATGTAAA 37

Qy       2930 GAAAAATTACATACATCCATGTTTGCTTACTTAACC 2965
             |||||||||||||||||||||||||||||||||||
Db          36 GAAAAATTACATACATCCATGTTTGCTTACTTAACC 1
```

```
RESULT 2
BI354756
LOCUS       BI354756                683 bp    mRNA    linear   EST 31-JUL-2001
DEFINITION  GM28157.5prime GM Drosophila melanogaster ovary pOT2 Drosophila
            melanogaster cDNA clone GM28157 5 similar to piwi: FBan0006122 GO:[
            ] located on: 2L 32C5-32C5;: 05/25/2001, mRNA sequence.
ACCESSION   BI354756
VERSION     BI354756.1  GI:15049198
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 683)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003630: arm:2L [10673952,10950527]
            estimated-cyto:32A5-32D2: 05/25/2001
            Plate: GM.281  row: E  column: 9
            High quality sequence stop: 586.
FEATURES             Location/Qualifiers
     source          1. .683
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GM28157"
                     /clone_lib="GM Drosophila melanogaster ovary pOT2"
                     /sex="female"
                     /dev_stage="newly eclosed females: germarium-stage 6"
                     /lab_host="XL1 Blue"
                     /note="Organ: ovary; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
BASE COUNT      197 a    145 c    160 g    181 t
ORIGIN

  Query Match             22.3%;  Score 678.2;  DB 13;  Length 683;
  Best Local Similarity   99.6%;  Pred. No. 1.8e-90;
  Matches  680;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        414 GAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAGAATCCTCGTGTC 473
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAGAATCCTCGTGTC 60

Qy        474 CGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTATCTATTCGACGGA 533
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTATCTATTCGACGGA 120

Qy        534 CTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTCAGCGGAAAGTCG 593
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTCAGCGGAAAGTCG 180

Qy        594 AAGCTGGACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATATCGTGTGCTGAG 653
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAGCTGGACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATATCGTGTGCTGAG 240

Qy        654 CCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCGATGAAGGGCCTAAATTTG 713
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCGATGAAGGGCCTAAATTTG 300

Qy        714 GAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATAAGGGAGTTCAAA 773
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATAAGGGAGTTCAAA 360

Qy        774 ATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAAGATATTTTATTG 833
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAAGATATTTTATTG 420

Qy        834 GGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGACATAATGCGACGT 893
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGACATAATGCGACGT 480

Qy        894 TGCTCACACAATCCGGCTCGTCATCAGGACGAAGTACGGGTAAATGTTTTGGACTTGATT 953
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TGCTCACACAATCCGGCTCGTCATCAGGACGAAGTACGGGTAAATGTTTTGGACTTGATT 540

Qy        954 GTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGATGTCGACTTTGGACAAACT 1013
                 |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        541 GTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGATGTCGACTGTGGACAAACT 600

Qy       1014 CCGAAATCAACATTCAGTTGCAAGGGTAGAGATATCAGTTTCGTGGAATACTATCTCACT 1073
                 |||||||||||||||| |||||||||||||||||||||||||||| ||||||||||||
Db        601 CCGAAATCAACATTCAGGTGCAAGGGTAGAGATATCAGTTTCGTGGGATACTATCTCACT 660

Qy       1074 AAATATAATATACGCATTCGCGA 1096
                 |||||||||||||||||||||||
Db        661 AAATATAATATACGCATTCGCGA 683


RESULT 3
AA802575
LOCUS       AA802575                620 bp    mRNA    linear   EST 23-APR-2001
DEFINITION  GM05853.5prime GM Drosophila melanogaster ovary BlueScript
            Drosophila melanogaster cDNA clone GM05853 5 similar to piwi:
            FBan0006122 located on: 2L 32C5-32C5;: 04/10/2001, mRNA sequence.
ACCESSION   AA802575
VERSION     AA802575.2  GI:13760741
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 620)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     On Feb 11, 1998 this sequence version replaced gi:2871694.
            Other_ESTs: GM05853.3prime
            Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003630: arm:2L [10673952,10950527]
            estimated-cyto:32A5-32D2: 04/10/2001
            Plate: GM.58  row: E  column: 5
            High quality sequence stop: 498
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .620
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GM05853"
                     /clone_lib="GM Drosophila melanogaster ovary BlueScript"
                     /sex="female"
                     /dev_stage="newly eclosed females: germarium-stage 6"
                     /lab_host="SOLR"
                     /note="Organ: ovary; Vector: BlueScript SK; Site_1: EcoRI;
                     Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      158 a    139 c    178 g    145 t
ORIGIN

  Query Match             20.2%;  Score 614.8;  DB 9;  Length 620;
  Best Local Similarity   99.4%;  Pred. No. 3.8e-81;
  Matches  616;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         44 TTGTGAAGTTTTAAGTGATCGCGAGTGCCAAAAAGTAACAATGGCTGATGATCAGGGACG 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTGTGAAGTTTTAAGTGATCGCGAGTGCCAAAAAGTAACAATGGCTGATGATCAGGGACG 60

Qy        104 TGGACGCAGGCGTCCANTTAACGAAGATGATTCCTCTACTTCCCGAGGTAGTGGTGATGG 163
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGACGCAGGCGTCCACTTAACGAAGATGATTCCTCTACTTCCCGAGGTAGTGGTGATGG 120

Qy        164 GCCGCGGGTGAAAGTATTCAGAGGATCTTCATCAGGTGACCCGAGAGCGGATCCTCGTAT 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCGCGGGTGAAAGTATTCAGAGGATCTTCATCAGGTGACCCGAGAGCGGATCCTCGTAT 180

Qy        224 AGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGAAGCTCCCAGGCGTGAAGGTGGCCCGCC 283
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGAGGCTTCAAGAGAGAGAGGAGCTCTCGAGGAAGCTCCCAGGCGTGAAGGTGGCCCGCC 240

Qy        284 AGAGCGAAAGCCGTGGGGTGACCAATATGATTACCTGAATACCCGTCCGGTTGAGCTGGT 343
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        241 AGAGCGAAAGCCGTGGGGTGACCAATATGATTACCTGGATACCCGTCCGGTTGAGCTGGT 300

Qy        344 ATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGANTAAA 403
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        301 ATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGATTAAA 360

Qy        404 AACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAGAA 463
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAGAA 420

Qy        464 TCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTATCT 523
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTATCT 480

Qy        524 ATTCGACGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTCAG 583
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATTCGACGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTCAG 540

Qy        584 CGGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATATC 643
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CGGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATATC 600

Qy        644 GTGTGCTGAGCCCCGCTTTT 663
              ||||||||||||||||||||
Db        601 GTGTGCTGAGCCCCGCTTTT 620


RESULT 4
BG637778/c
LOCUS       BG637778                589 bp    mRNA    linear   EST 23-APR-2001
DEFINITION  LD03429.3prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD03429 3 similar to piwi:
            FBan0006122 located on: 2L 32C5-32C5;: 04/10/2001, mRNA sequence.
ACCESSION   BG637778
VERSION     BG637778.1  GI:13765820
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 589)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: LD03429.5prime
            Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
```

```
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Based upon the presence of a XhoI site followed by a run of 14 or
            more T residues at the beginning of the sequence, this clone was
            polyadenylated. The resulting Poly-T sequence has been removed. hit
            genomic AE003630: arm:2L [10673952,10950527]
            estimated-cyto:32A5-32D2: 04/10/2001
            Plate: LD.34  row: C  column: 5
            High quality sequence stop: 446.
FEATURES             Location/Qualifiers
     source          1. .589
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LD03429"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1: EcoRI
                     ; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      173 a    105 c    128 g    182 t      1 others
ORIGIN

  Query Match           19.2%;  Score 583.8;  DB 12;  Length 589;
  Best Local Similarity  99.3%;  Pred. No. 1.4e-76;
  Matches  585;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2268 TCCATGAACACGCGCTTCTTCCTCAACGGACAAAATCCTCCGCCTGGTACTATAGTTGAT 2327
             ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db        589 TCCATGAACACGCGCTTCTTCCTCAACGGACAAAATCCTCCGCCTGTNACTATAGTTGAT 530

Qy       2328 GACGTTATAACTCTGCCCGAGAGATACGACTTTTATCTGGTCTCGCAACAAGTTCGTCAG 2387
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        529 GACGTTATAACTCTGCCCGAGAGATACGACTTTTATCTGGTCTCGCAACAAGTTCGTCAG 470

Qy       2388 GGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAGCAGCATGGGTNTCTCACCGGAG 2447
             |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        469 GGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAGCAGCATGGGTCTCTCACCGGAG 410

Qy       2448 AAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTACTACAATTGGTCGGGCACCACA 2507
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        409 AAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTACTACAATTGGTCGGGCACCACA 350

Qy       2508 CGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGCTACCCTCGTGGGTACGAACTTG 2567
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        349 CGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGCTACCCTCGTGGGTACGAACTTG 290

Qy       2568 CACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTATTATCTATAATTGGATATAATT 2627
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        289 CACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTATTATCTATAATTGGATATAATT 230

Qy       2628 TAGAATGGAGTATTAATCCTTACTAAGAGGCCATATATGAAACTAGCCCAGACATTTATA 2687
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        229 TAGAATGGAGTATTAATCCTTACTAAGAGGCCATATATGAAACTAGCCCAGACATTTATA 170

Qy       2688 CTTTTTCAATACTTCCTTACTTTTGCTAAGCACTTCAGCATTTATGACTAAATATTTTGT 2747
             |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        169 CTTTTTCAATACTTTCTTACTTTTGCTAAGCACTTCAGCATTTATGACTAAATATTTTGT 110

Qy       2748 ATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGCAAAATTGAGGATTAAGATTCTG 2807
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        109 ATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGCAAAATTGAGGATTAAGATTCTG 50

Qy       2808 GTATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAACATTTAAAGTT 2856
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db          49 GTATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAACATTTAAAGTT 1


RESULT 5
AA202789
LOCUS       AA202789                 567 bp    mRNA    linear   EST 23-APR-2001
DEFINITION  LD03429.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD03429 5 similar to piwi:
            FBan0006122 located on: 2L 32C5-32C5;: 04/10/2001, mRNA sequence.
ACCESSION   AA202789
VERSION     AA202789.2  GI:13765821
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 567)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:2776690.
            Other_ESTs: LD03429.3prime
            Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003630: arm:2L [10673952,10950527]
            estimated-cyto:32A5-32D2: 04/10/2001
            Plate: LD.34  row: C  column: 5
            High quality sequence stop: 524
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .567
                     /organism="Drosophila melanogaster"
                     /db_xref="BDGP_EST:BDcln002872"
                     /db_xref="taxon:7227"
                     /clone="LD03429"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1: EcoRI
                     ; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      150 a    147 c    130 g    140 t
ORIGIN

  Query Match           18.6%;  Score 566;  DB 9;  Length 567;
  Best Local Similarity  99.8%;  Pred. No. 5.8e-74;
  Matches  566;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1876 CCTGGATGATCGAACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGA 1935
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCTGGATGATCGAACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGA 60

Qy       1936 GCACACGAGATCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAA 1995
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCACACGAGATCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAA 120

Qy       1996 ACTCCACGTACTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCC 2055
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 ACTCCACGTACTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCC 180

Qy       2056 TTTGGCCGATGATAGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCAT 2115
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 TTTGGCCGATGATAGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCAT 240

Qy       2116 CTCGAATCGTATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAAT 2175
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 CTCGAATCGTATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAAT 300

Qy       2176 TTGAAGTCAAGGACATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCC 2235
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 TTGAAGTCAAGGACATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCC 360
```

```
Qy       2236 CACCGCAATTAGCTTATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACG 2295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CACCGCAATTAGCTTATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACG 420

Qy       2296 GACAAAATCCTCCGCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACG 2355
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GACAAAATCCTCCGCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACG 480

Qy       2356 ACTTTTATCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATG 2415
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ACTTTTATCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATG 540

Qy       2416 TTCTTTATAGCAGCATGGGTNTCTCAC 2442
              |||||||||||||||||||| ||||||
Db        541 TTCTTTATAGCAGCATGGGTCTCTCAC 567
```

RESULT 6
BI214891

LOCUS	BI214891 573 bp mRNA linear EST 11-JUL-2001
DEFINITION	RE21038.5prime RE Drosophila melanogaster normalized Embryo pFlc-1 Drosophila melanogaster cDNA clone RE21038 5 similar to piwi: FBan0006122 located on: 2L 32C5-32C5;: 04/12/2001, mRNA sequence.
ACCESSION	BI214891
VERSION	BI214891.1 GI:14692615
KEYWORDS	EST.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 573)
AUTHORS	Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin ,G.M.
TITLE	BDGP/HHMI RE Drosophila EST Project
JOURNAL	Unpublished (2001)
COMMENT	Contact: Stapleton, M. BDGP Lawrence Berkeley National Lab One Cyclotron Rd, Berkeley, CA 94720, USA Fax: 510 486 6798 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu hit genomic AE003630: arm:2L [10673952,10950527] estimated-cyto:32A5-32D2: 04/12/2001 Plate: RE.210 row: D column: 2 High quality sequence stop: 487.

FEATURES Location/Qualifiers

```
     source          1. .573
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RE21038"
                     /clone_lib="RE Drosophila melanogaster normalized Embryo
                     pFlc-1"
                     /sex="male and female"
                     /dev_stage="0-24 hours mixed stage embryonic"
                     /lab_host="DH5-alpha TonA"
                     /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
                     BamHI; Library was kindly generated by Piero Carninci at
                     the RIKEN. The library was normalized and excised using
                     Cre recombinase. Plasmid cDNA library."
BASE COUNT      149 a    129 c    161 g    134 t
ORIGIN
```

```
  Query Match             18.4%;  Score 559.4;  DB 13;  Length 573;
  Best Local Similarity   98.6%;  Pred. No. 5.4e-73;
  Matches  563;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 CTGAGTCCAAAGCGTCGTTTTCAAAGTACTCTTTCAGTTTCCATTGTGAAGTTTTAAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          3 CTGAGTCCAAAGCGTCGTTTTCAAAGTACTCTTTCAGTTTCCATTGTGAAGTTTTAAGTG 62

Qy         61 ATCGCGAGTGCCAAAAAGTAACAATGGCTGATGATCAGGGACGTGGACGCAGGCGTCCAN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 ATCGCGAGTGCCAAAAAGTAACAATGGCTGATGATCAGGGACGTGGACGCAGGCGTCCAC 122

Qy        121 TTAACGAAGATGATTCCTCTACTTCCCGAGGTAGTGGTGATGGGCCGCGGGTGAAAGTAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 TTAACGAAGATGATTCCTCTACTTCCCGAGGTAGTGGTGATGGGCCGCGGGTGAAAGTAT 182

Qy        181 TCAGAGGATCTTCATCAGGTGACCCGAGAGCGGATCCTCGTATAGAGGCTTCAAGAGAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 TCAGAGGATCTTCATCAGGTGACCCGAGAGCGGATCCTCGTATAGAGGCTTCAAGAGAGA 242

Qy        241 GAAGAGCTCTCGAGGAAGCTCCCAGGCGTGAAGGTGGCCCGCCAGAGCGAAAGCCGTGGG 300
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        243 GAAGAGCTCTCGAGGAAGCTCCCAGGCGTGAAGGTGGCCCGACAGAGCGAAAGCCGTGGG 302

Qy        301 GTGACCAATATGATTACCTGAATACCCGTCCGGTTGAGCTGGTATCCAAGAAGGGAACCG 360
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        303 GTGACCAATATGATTACCTGAATACCCGTCCGGCTGAGCTGGTATCCAAGAAGGGAACCG 362

Qy        361 ATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGANTAAAAACCAAGCCGGAATGGC 420
              |||||||||||||||||||||||||||||||| ||| ||||||||||||||||||||||
Db        363 ATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTCCGATTAAAAACCAAGCCGGAATGGC 422

Qy        421 GGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAGAATCCTCGTGTCCGTATGG 480
              |||||||||||||||||||||||||||||| ||||| ||||||||||||||||||||||
Db        423 GGATCGTTCATTATCACGTGGAGTTTGAGCCGAGCATCGAGAATCCTCGTGTCCGTATGG 482

Qy        481 GAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTATCTATTCGACGGACTGCAAC 540
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        483 GAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTATCTATTCGATGGACTGCAAC 542

Qy        541 TGTTCACCACCAGGAAATTCGAGCAGGAAAT 571
              |||||||||||||||||||||||||||||||
Db        543 TGTTCACCACCAGGAAATTCGAGCAGGAAAT 573
```

RESULT 7
BI372643

LOCUS	BI372643 538 bp mRNA linear EST 01-AUG-2001
DEFINITION	RE59653.5prime RE Drosophila melanogaster normalized Embryo pFlc-1 Drosophila melanogaster cDNA clone RE59653 5 similar to piwi: FBan0006122 GO:[] located on: 2L 32C5-32C5;: 05/16/2001, mRNA sequence.
ACCESSION	BI372643
VERSION	BI372643.1 GI:15068671
KEYWORDS	EST.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 538)
AUTHORS	Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin ,G.M.
TITLE	BDGP/HHMI RE Drosophila EST Project
JOURNAL	Unpublished (2001)
COMMENT	Contact: Stapleton, M. BDGP Lawrence Berkeley National Lab One Cyclotron Rd, Berkeley, CA 94720, USA Fax: 510 486 6798 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu hit genomic AE003630: arm:2L [10673952,10950527] estimated-cyto:32A5-32D2: 05/16/2001 Plate: RE.596 row: E column: 5 High quality sequence stop: 517.

```
FEATURES             Location/Qualifiers
     source          1. .538
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RE59653"
                     /clone_lib="RE Drosophila melanogaster normalized Embryo
                     pFlc-1"
                     /sex="male and female"
                     /dev_stage="0-24 hours mixed stage embryonic"
                     /lab_host="DH5-alpha TonA"
                     /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
                     BamHI; Library was kindly generated by Piero Carninci at
                     the RIKEN. The library  was normalized and excised using
                     Cre recombinase. Plasmid cDNA library."
BASE COUNT      136 a    122 c    153 g    127 t
ORIGIN

  Query Match           17.1%;  Score 521.6;  DB 13;  Length 538;
  Best Local Similarity   98.0%;  Pred. No. 2e-67;
  Matches  527;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy         17 GTTTTCAAAGTACTCTTTCAGTTTCCATTGTGAAGTTTTAAGTGATCGCGAGTGCCAAAA 76
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTTTTCAAAGTACTCTTTCAGTTTCCATTGTGAAGTTTTAAGTGATCGCGAGTGCCAAAA 60

Qy         77 AGTAACAATGGCTGATGATCAGGGACGTGGACGCAGGCGTCCANTTAACGAAGATGATTC 136
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db         61 AGTAACAATGGCTGATGATCAGGGACGTGGACGCAGGCGTCCACTTAACGAAGATGATTC 120

Qy        137 CTCTACTTCCCGAGGTAGTGGTGATGGGCCGCGGGTGAAAGTATTCAGAGGATCTTCATC 196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTCTACTTCCCGAGGTAGTGGTGATGGGCCGCGGGTGAAAGTATTCAGAGGATCTTCATC 180

Qy        197 AGGTGACCCGAGAGCGGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGA 256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGGTGACCCGAGAGCGGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGA 240

Qy        257 AGCTCCCAGGCGTGAAGGTGGCCCGCCAGAGCGAAAGCCGTGGGGTGACCAATATGATTA 316
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        241 AGCTCCCAGGCGTGAAGGTGGCCCGACAGAGCGAAAGCCGTGGGGTGACCAATATGATTA 300

Qy        317 CCTGAATACCCGTCCGGTTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTCCCGGTCAT 376
              |||||||||||||||| |||||||||||||||||||| ||||||||||||||||||||||
Db        301 CCTGAATACCCGTCCGGCTGAGCTGGTATCCAAGAAGGGAACCAATGGCGTCCCGGTCAT 360

Qy        377 GCTGCAGACGAACTTTTTTCGANTAAAAACCAAGCCGGAATGGCGGATCGTTCATTATCA 436
              |||||||||||||||||| ||| ||||||||||||||||||||||||||||||||||||
Db        361 GCTGCAGACGAACTTTTTCCGATTAAAAACCAAGCCGGAATGGCGGATCGTTCATTATCA 420

Qy        437 CGTGGAGTTTGTGCCGACCATCGAGAATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCA 496
              |||||||||| |||| ||||||||||||||||||||||||||||||||||||||||||||
Db        421 CGTGGAGTTTGAGCCGAGCATCGAGAATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCA 480

Qy        497 TGCTAACCTTCTGGGATCAGGCTATCTATTCGACGGACTGCAACTGTTCACCACCAGG 554
              |||||||||||||||||||||||||||||||| ||||||  ||||||||||||||||
Db        481 TGCTAACCTTCTGGGATCAGGCTATCTATTCGATGGACTGGCACTGTTCACCACCAGG 538


RESULT 8
AI061964/c
LOCUS       AI061964                500 bp    mRNA    linear   EST 19-APR-2001
DEFINITION  LD35072.3prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD35072 3prime, mRNA sequence.
ACCESSION   AI061964
VERSION     AI061964.1  GI:3337803
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Based upon the presence of a XhoI site followed by a run of 14 or
            more T residues at the beginning of the sequence, this clone
            probably contains an inverted insert. The sequence has been trimmed
            and the T residues removed.
            Plate: 350  row: F  column: 12
            High quality sequence stop: 437.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LD35072"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
BASE COUNT      170 a     76 c     88 g    166 t
ORIGIN

  Query Match           16.3%;  Score 496.8;  DB 9;  Length 500;
  Best Local Similarity   99.6%;  Pred. No. 9e-64;
  Matches  498;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2466 TACAAGATGTGCCACTTGTACTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGC 2525
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 TACAAGATGGGCCACTTGTACTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGC 441

Qy       2526 CAGTACGCTAAGAAGCTAGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAAC 2585
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 CAGTACGCTAAGAAGCTAGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAAC 381

Qy       2586 GCGCTCGAAAAGAAGTTTTATTATCTATAATTGGATATAATTTAGAATGGAGTATTAATC 2645
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 GCGCTCGAAAAGAAGTTTTATTATCTATAATTGGATATAATTTAGAATGGAGTATTAATC 321

Qy       2646 CTTACTAAGAGGCCATATATGAAACTAGCCCAGACATTTATACTTTTTCAATACTTCCTT 2705
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        320 CTTACTAAGAGGCCATATATGAAACTAGCCCAGACATTTATACTTTTTCAATACTTTCTT 261

Qy       2706 ACTTTTGCTAAGCACTTCAGCATTTATGACTAAATATTTTGTATTTGAAATGCATTACTG 2765
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 ACTTTTGCTAAGCACTTCAGCATTTATGACTAAATATTTTGTATTTGAAATGCATTACTG 201

Qy       2766 CTCTTTTTTCAAACAAAAGCAAAATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACC 2825
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 CTCTTTTTTCAAACAAAAGCAAAATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACC 141

Qy       2826 AGAGGAAATTCCCAAACAAACATTTAAAGTTATCTATCAAGACATGTTCATTAATTTGGA 2885
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 AGAGGAAATTCCCAAACAAACATTTAAAGTTATCTATCAAGACATGTTCATTAATTTGGA 81

Qy       2886 ATATAATTACTTTATTTTTTATTGTATATTTTAGTTTATGTAAAGAAAAATTACATACAT 2945
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         80 ATATAATTACTTTATTTTTTATTGTATATTTTAGTTTATGTAAAGAAAAATTACATACAT 21

Qy       2946 CCATGTTTGCTTACTTAACC 2965
              ||||||||||||||||||||
Db         20 CCATGTTTGCTTACTTAACC 1


RESULT 9
BI241150
```

```
LOCUS       BI241150                498 bp    mRNA    linear   EST 12-JUL-2001
DEFINITION  RE38167.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
            Drosophila melanogaster cDNA clone RE38167 5 similar to piwi:
            FBan0006122 located on: 2L 32C5-32C5;: 05/12/2001, mRNA sequence.
ACCESSION   BI241150
VERSION     BI241150.1  GI:14709800
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
            ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
            ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
            Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
            ,G.M.
  TITLE     BDGP/HHMI RE Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003630: arm:2L [10673952,10950527]
            estimated-cyto:32A5-32D2: 05/12/2001
            Plate: RE.381  row: F  column: 7
            High quality sequence stop: 442.
FEATURES             Location/Qualifiers
     source          1. .498
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RE38167"
                     /clone_lib="RE Drosophila melanogaster normalized Embryo
                     pFlc-1"
                     /sex="male and female"
                     /dev_stage="0-24 hours mixed stage embryonic"
                     /lab_host="DH5-alpha TonA"
                     /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
                     BamHI; Library was kindly generated by Piero Carninci at
                     the RIKEN. The library  was normalized and excised using
                     Cre recombinase. Plasmid cDNA library."
BASE COUNT      127 a    111 c    143 g    117 t
ORIGIN

  Query Match            16.0%;  Score 486.4;  DB 13;  Length 498;
  Best Local Similarity   98.4%;  Pred. No. 3e-62;
  Matches  490;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy         17 GTTTTCAAAGTACTCTTTCAGTTTCCATTGTGAAGTTTTAAGTGATCGCGAGTGCCAAAA 76
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTTTTCAAAGTACTCTTTCAGTTTCCATTGTGAAGTTTTAAGTGATCGCGAGTGCCAAAA 60

Qy         77 AGTAACAATGGCTGATGATCAGGGACGTGGACGCAGGCGTCCANTTAACGAAGATGATTC 136
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db         61 AGTAACAATGGCTGATGATCAGGGACGTGGACGCAGGCGTCCACTTAACGAAGATGATTC 120

Qy        137 CTCTACTTCCCGAGGTAGTGGTGATGGGCCGCGGGTGAAAGTATTCAGAGGATCTTCATC 196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTCTACTTCCCGAGGTAGTGGTGATGGGCCGCGGGTGAAAGTATTCAGAGGATCTTCATC 180

Qy        197 AGGTGACCCGAGAGCGGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGA 256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGGTGACCCGAGAGCGGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGA 240

Qy        257 AGCTCCCAGGCGTGAAGGTGGCCCGCCAGAGCGAAAGCCGTGGGGTGACCAATATGATTA 316
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        241 AGCTCCCAGGCGTGAAGGTGGCCCGACAGAGCGAAAGCCGTGGGGTGACCAATATGATTA 300

Qy        317 CCTGAATACCCGTCCGGTTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTCCCGGTCAT 376
              |||||||||||||||| |||||||||||||||||||||||||||||||| ||||||||||||
Db        301 CCTGAATACCCGTCCGGCTGAGCTGGTATCCAAGAAGGGAACCAATGGCGTCCCGGTCAT 360

Qy        377 GCTGCAGACGAACTTTTTTCGANTAAAAACCAAGCCGGAATGGCGGATCGTTCATTATCA 436
              |||||||||||||||||| ||| ||||||||||||||||||||||||||||||||||||
Db        361 GCTGCAGACGAACTTTTTCCGATTAAAAACCAAGCCGGAATGGCGGATCGTTCATTATCA 420

Qy        437 CGTGGAGTTTGTGCCGACCATCGAGAATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCA 496
              |||||||||| ||||| ||||||||||||||||||||||||||||||||||||||||||
Db        421 CGTGGAGTTTGAGCCGAGCATCGAGAATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCA 480

Qy        497 TGCTAACCTTCTGGGATC 514
              ||||||||||||||||||
Db        481 TGCTAACCTTCTGGGATC 498


RESULT 10
BG637781/c
LOCUS       BG637781                489 bp    mRNA    linear   EST 23-APR-2001
DEFINITION  LD03506.3prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD03506 3 similar to piwi:
            FBan0006122 located on: 2L 32C5-32C5;: 04/10/2001, mRNA sequence.
ACCESSION   BG637781
VERSION     BG637781.1  GI:13765829
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 489)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: LD03506.5prime
            Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Based upon the presence of a XhoI site followed by a run of 14 or
            more T residues at the beginning of the sequence, this clone was
            polyadenylated. The resulting Poly-T sequence has been removed. hit
            genomic AE003630: arm:2L [10673952,10950527]
            estimated-cyto:32A5-32D2: 04/10/2001
            Plate: LD.35  row: A  column: 6
            High quality sequence stop: 422.
FEATURES             Location/Qualifiers
     source          1. .489
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LD03506"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1: EcoRI
                     ; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      143 a     88 c     98 g    160 t
ORIGIN

  Query Match            15.4%;  Score 468.8;  DB 12;  Length 489;
  Best Local Similarity   97.3%;  Pred. No. 1.2e-59;
  Matches  476;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       2368 TCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAGCA 2427
              ||||||||||||| |||||||||||||||||| || ||||||||||||||||||
Db        489 TCTCGCAACAAGTTAGTCAGGGTACAGTGTCGCCGACAAGATACAATGTTCTTTATAGCA 430
```

```
Qy       2428 GCATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTACT 2487
              |||||||| ||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        429 GCATGGGTCTCTAACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTACT 370

Qy       2488 ACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGCTA 2547
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        369 ACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGCAGCTAGCTA 310

Qy       2548 CCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTATT 2607
              |||||||||||||||||||||| |||||||||||||||||||||||||||||||| |
Db        309 CCCTCGTGGGTACGAACTTGCACTATATTCCGCAAAACGCGCTCGAAAAGAAGTTTTCTC 250

Qy       2608 ATCTATAATTGGATATAATTTAGAATGGAGTATTAATCCTTACTAAGAGGCCATATATGA 2667
              |||||||||||||||||||||||||||||||||| || ||||||||||||||||||||
Db        249 ATCTATAATTGGATATAATTTAGAATGGAGTATTAATCCGTATTAAGAGGCCATATATGA 190

Qy       2668 AACTAGCCCAGACATTTATACTTTTTCAATACTTCCTTACTTTTGCTAAGCACTTCAGCA 2727
              |||||||||||||||||||||||||||||||| ||||||| ||||||||||||||||||
Db        189 AACTAGCCCAGACATTTATACTTTTTCAATACTTTCTTACTTCTGCTAAGCACTTCAGCA 130

Qy       2728 TTTATGACTAAATATTTTGTATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGCAA 2787
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 TTTATGACTAAATATTTTGTATTTGAAATGCATTACTGCTCTTTTTTCAAACAAAAGCAA 70

Qy       2788 AATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAACA 2847
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 AATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCCAAACAAACA 10

Qy       2848 TTTAAAGTT 2856
              |||||||||
Db          9 TTTAAAGTT 1
```

RESULT 11
BI168498
LOCUS BI168498 433 bp mRNA linear EST 09-JUL-2001
DEFINITION RE09253.5prime RE Drosophila melanogaster normalized Embryo pFlc-1 Drosophila melanogaster cDNA clone RE09253 5 similar to piwi: FBan0006122 located on: 2L 32C5-32C5;: 04/11/2001, mRNA sequence.
ACCESSION BI168498
VERSION BI168498.1 GI:14634305
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 433)
AUTHORS Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin,G.M.
TITLE BDGP/HHMI RE Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AE003630: arm:2L [10673952,10950527]
estimated-cyto:32A5-32D2: 04/11/2001
Plate: RE.92 row: E column: 5
High quality sequence stop: 356.
FEATURES Location/Qualifiers
source 1. .433
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="RE09253"
/clone_lib="RE Drosophila melanogaster normalized Embryo pFlc-1"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha TonA"
/note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."
BASE COUNT 112 a 91 c 127 g 102 t 1 others
ORIGIN

Query Match 13.4%; Score 409.4; DB 13; Length 433;
Best Local Similarity 97.7%; Pred. No. 6.6e-51;
Matches 424; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

```
Qy         14 GTCGTTTTCAAAGTACTCTTTCAGTTTCCATTGTGAAGTTTTAAGTGATCGCGAGTGCCA 73
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db          1 GTCGTTTTCAAAGTACTCTTTCAGTTTCCATTGTGAAGTTTTAAGTGATCACGAGTGCCA 60

Qy         74 AAAAGTAACAATGGCTGATGATCAGGGACGTGGACGCAGGCGTCCANTTAACGAAGATGA 133
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db         61 AAAAGTAACAATGGCTGATGATCAGGGACGTGGACGCAGGCGTCCACTTAACGAAGATGA 120

Qy        134 TTCCTCTACTTCCCGAGGTAGTGGTGATGGGCCGCGGGTGAAAGTATTCAGAGGATCTTC 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTCCTCTACTTCCCGAGGTAGTGGTGATGGGCCGCGGGTGAAAGTATTCAGAGGATCTTC 180

Qy        194 ATCAGGTGACCCGAGAGCGGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCAGGTGACCCGAGAGCGGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGA 240

Qy        254 GGAAGCTCCCAGGCGTGAAGGTGGCCCGCCAGAGCGAAAGCCGTGGGGTGACCAATATGA 313
              |||||||||||||||||||||| ||| |||||||||| ||||||||||||||||||||
Db        241 GGAAGCTCCCAGGCGTGAAGGTGGTCCGACAGAGCGAAAGTCGTGGGGTGACCAATATGA 300

Qy        314 TTACCTGAATACCCGTCCGGTTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTCCCGGT 373
              |||||||||||||||||||| ||||||||||||||||||||||||||||||| |||||
Db        301 TTACCTGAATACCCGTCCGGCTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTTCCGGT 360

Qy        374 CATGCTGCAGACGAACTTTTTTCGANTAAAAACCAAGCCGGAATGGCGGATCGTTCATTA 433
              |||||||||||||||||||| ||| | ||||||||||||||||||||||||||||||||
Db        361 CATGCTGCAGACGAACTTTTTNCGATT-AAAACCAAGCCGGAATGGCGGATCGTTCATTA 419

Qy        434 TCACGTGGAGTTTG 447
              ||||||||||||||
Db        420 TCACGTGGAGTTTG 433
```

RESULT 12
CNS008NM
LOCUS CNS008NM 988 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR17K08 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL052102
VERSION AL052102.1 GI:4933954
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 988)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
     source          1. .988
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR17K08"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      294 a    203 c    185 g    272 t     34 others
ORIGIN
```

Query Match 13.0%; Score 395.6; DB 17; Length 988;
Best Local Similarity 96.2%; Pred. No. 4.3e-49;
Matches 429; Conservative 7; Mismatches 6; Indels 4; Gaps 3;

```
Qy       1731 AGATACTCATCAATCAAAAAGAGAGGATACGTTGACAGGGCGGTGCCAACTCAAGTTGTG 1790
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        406 AGATACTCATCAATCAAAAAGAGAGGATACGTTGACAGGGCGGTGCCAACTCAAGTTGTG 465

Qy       1791 ACCCTTAAAACGACCAAGAACCGTAGCCTTATGAGCATTGCCACCAAAATAGCAATCCAA 1850
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        466 ACCCTTAAAACGACCAAGAACCGTAGCCTTATGAGCATTGCCACCAAAATAGCAATCCAA 525

Qy       1851 CTGAATTGCAAGTTGGGATATACACCCTGGATGATCGAACTACCCTTGTCCGGACTGATG 1910
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        526 CTGAATTGCAAGTTGGGATATACACCCTGGATGATCGAACTACCCTTGTCCGGACTGATG 585

Qy       1911 ACAATTGGCTTTGACATTGCGAAGAGCACACGAGATCGGAAGAGGGCCTACGGAGCATTG 1970
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        586 ACAATTGGCTTTGACATTGCGAAGAGCACACGAGATCGGAAGAGGGCCTACGGAGCATTG 645

Qy       1971 ATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTTCAGCACAGTCACGGAGTGCAGC 2030
             |||||||||||||||||||||||||||||||||||||||||||||||||||| :||||||
Db        646 ATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTTCAGCACAGTCACGGAG-KCAGC 704

Qy       2031 GCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATGATAGCAAAGGCCCTGCGCCAATAT 2090
             ||||||||||||||||||||||||||||:|:||||||||||||||||||||||||||||||
Db        705 GCCTTTGATGTGCTCGCTAACACCCTTTKGSCGATGATAGCAAAGGCCCTGCGCCAATAT 764

Qy       2091 CAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATTTTATCGAGACGGTGTGAGCTCC 2150
             |||||||||||||||||||||||||||||||| |||||||||||||||||||:||
Db        765 CAACATGAGCATAGGAAGCTGCCATCTCGAATCGTA-TTTATCGAGACGGTGTGAGCYCC 823

Qy       2151 GGCTCTCTAAAGCAGCTTTTTGAATT 2176
             |     :  ||:||:|||||||| ||
Db        824 G--CTCYCTAARCARCTTTTTGATTT 847
```

```
RESULT 13
BI163833
LOCUS       BI163833                385 bp    mRNA    linear   EST 09-JUL-2001
DEFINITION  RE03347.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
            Drosophila melanogaster cDNA clone RE03347 5 similar to piwi:
            FBan0006122 located on: 2L 32C5-32C5;: 04/11/2001, mRNA sequence.
ACCESSION   BI163833
VERSION     BI163833.1  GI:14629639
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 385)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
            ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
            ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
            Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
            ,G.M.
  TITLE     BDGP/HHMI RE Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003630: arm:2L [10673952,10950527]
            estimated-cyto:32A5-32D2: 04/11/2001
            Plate: RE.33  row: D  column: 11
            High quality sequence stop: 366.
FEATURES             Location/Qualifiers
     source          1. .385
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RE03347"
                     /clone_lib="RE Drosophila melanogaster normalized Embryo
                     pFlc-1"
                     /sex="male and female"
                     /dev_stage="0-24 hours mixed stage embryonic"
                     /lab_host="DH5-alpha TonA"
                     /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
                     BamHI; Library was kindly generated by Piero Carninci at
                     the RIKEN. The library was normalized and excised using
                     Cre recombinase. Plasmid cDNA library."
BASE COUNT      100 a     85 c    113 g     86 t      1 others
ORIGIN
```

Query Match 12.3%; Score 375.6; DB 13; Length 385;
Best Local Similarity 98.2%; Pred. No. 6.5e-46;
Matches 378; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy         17 GTTTTCAAAGTACTCTTTCAGTTTCCATTGTGAAGTTTTAAGTGATCGCGAGTGCCAAAA 76
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTTTTCAAAGTACTCTTTCAGTTTCCATTGTGAAGTTTTAAGTGATCGCGAGTGCCAAAA 60

Qy         77 AGTAACAATGGCTGATGATCAGGGACGTGGACGCAGGCGTCCANTTAACGAAGATGATTC 136
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db         61 AGTAACAATGGCTGATGATCAGGGACGTGGACGCAGGCGTCCACTTAACGAAGATGATTC 120

Qy        137 CTCTACTTCCCGAGGTAGTGGTGATGGGCCGCGGGTGAAAGTATTCAGAGGATCTTCATC 196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTCTACTTCCCGAGGTAGTGGTGATGGGCCGCGGGTGAAAGTATTCAGAGGATCTTCATC 180

Qy        197 AGGTGACCCGAGAGCGGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGA 256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGGTGACCCGAGAGCGGATCCTCGTATAGAGGCTTCAAGAGAGAGAAGAGCTCTCGAGGA 240

Qy        257 AGCTCCCAGGCGTGAAGGTGGCCCGCCAGAGCGAAAGCCGTGGGGTGACCAATATGATTA 316
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        241 AGCTCCCAGGCGTGAAGGTGGCCCGACAGAGCGAAAGCCGTGGGGTGACCAATATGATTA 300

Qy        317 CCTGAATACCCGTCCGGTTGAGCTGGTATCCAAGAAGGGAACCGATGGCGTCCCGGTCAT 376
              |||||||||||||||| |||||||||||||||||||||||| ||||||||||||||||
Db        301 CCTGAATACCCGTCCGGCTGAGCTGGTATCCAAGAAGGGAACCAATGGCGTCCCGGTCAT 360

Qy        377 GCTGCAGACGAACTTTTTTCGANTA 401
              |||||||||||||||| | ||| ||
Db        361 GCTGCAGACGAACTTTNTCCGATTA 385
```

```
RESULT 14
AW941950/c
LOCUS       AW941950                410 bp    mRNA    linear   EST 23-APR-2001
DEFINITION  GM05853.3prime GM Drosophila melanogaster ovary BlueScript
            Drosophila melanogaster cDNA clone GM05853 3 similar to AF104354:
            Drosophila melanogaster PIWI (piwi) mRNA, complete cds, mRNA
            sequence.
```

```
ACCESSION   AW941950
VERSION     AW941950.1  GI:8119649
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 410)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: GM05853.5prime
            Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Based upon the presence of a XhoI site followed by a run of 14 or
            more T residues at the beginning of the sequence, this clone was
            polyadenylated. The resulting Poly-T sequence has been removed. hit
            genomic AE003630: Drosophila melanogaster genomic scaffold
            142000013386055 section 23 of 63, complete sequence.: 02/20/2001
            Plate: GM.58  row: E  column: 5
            High quality sequence stop: 255.
FEATURES             Location/Qualifiers
     source          1. .410
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GM05853"
                     /clone_lib="GM Drosophila melanogaster ovary BlueScript"
                     /sex="female"
                     /dev_stage="newly eclosed females: germarium-stage 6"
                     /lab_host="SOLR"
                     /note="Organ: ovary; Vector: BlueScript SK; Site_1: EcoRI;
                     Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      159 a     52 c     62 g    137 t
ORIGIN

  Query Match             10.7%;  Score 325.2;  DB 10;  Length 410;
  Best Local Similarity   90.2%;  Pred. No. 1.6e-38;
  Matches  348;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       2659 CATATATGAAACTAGCCCAGACATTTATACTTTTTCAATACTTCCTTACTTTTGCTAAGC 2718
              | | ||||||||| |||||||||||||||||||||||||||||| ||||||||||||||||
Db        410 CTTTTATGAAACTGGCCCAGACATTTATACTTTTTCAATACTTTCTTACTTTTGCTAAGC 351

Qy       2719 ACTTCAGCATTTATGACTAAATATTTTGTATTTGAAATGCATTACTGCTCTTTTTTCAAA 2778
              |||||||||||||||| |||||||||||||| |||||||||||||||||||||||||||||
Db        350 ACTTCAGCATTTATGGCTAAATATTTTGTAATTGAAATGCATTACTGCTCTTTTTTCAAA 291

Qy       2779 CAAAAGCAAAATTGAGGATTAAGATTCTGGTATTTAAGCATAAGACCAGAGGAAATTCCC 2838
              |||| |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        290 CAAAGGCAAAATTGAGGATTAAGATTCTGGTATTTAAGCATAAGCCCAGAGGAAATTCCC 231

Qy       2839 AAACAAACATTTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTT 2898
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        230 AAACAAACATTTAAAGTTATCTATCAAGACATGTTCATTAATTTGGAATATAATTACTTT 171

Qy       2899 ATTTTTTATTGTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTA 2958
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        170 ATTTTTTAGTGTATATTTTAGTTTATGTAAAGAAAAATTACATACATCCATGTTTGCTTA 111

Qy       2959 CTTAACCACACATTCATGGCTGCTTATATTCGTGAAAAAAAAAAAAAAAAAAAAAAAAAA 3018
              |||||||||||||||||||||||||||||||||||| | |    |  ||  | | |
Db        110 CTTAACCACACATTCATGGCTGCTTATATTCGTGAAACATATTTTATTAATTATACAGGT 51

Qy       3019 AAAAAAAAAAAAAAAAAAAAAAAAA 3044
              |   ||     | ||  | || | ||
```

```
Db          50 ATTTAACTTTCACAAGGATAATACAA 25


RESULT 15
AA817640
LOCUS       AA817640                775 bp    mRNA    linear   EST 19-APR-2001
DEFINITION  LD24150.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD24150 5prime, mRNA sequence.
ACCESSION   AA817640
VERSION     AA817640.1  GI:2887249
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 775)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 241  row: E  column: 2
            High quality sequence stop: 441.
FEATURES             Location/Qualifiers
     source          1. .775
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LD24150"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
BASE COUNT      194 a    204 c    187 g    190 t
ORIGIN

  Query Match             6.5%;  Score 196.8;  DB 9;  Length 775;
  Best Local Similarity   55.5%;  Pred. No. 7.6e-20;
  Matches  400;  Conservative    0;  Mismatches  318;  Indels    3;  Gaps    1;

Qy       1827 ATTGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGATC 1886
              || ||||| ||  | |  || ||| ||||   ||| |||       | ||||||  | |
Db          4 ATCGCCACAAAGGTGGTTATTCAAATGAACGCCAAATTGATGGGAGCTCCCTGGCAGGTA 63

Qy       1887 GAACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGAT 1946
              |   | ||| |   ||| ||||||||  |||| || ||  |     |     |     ||
Db         64 GTGATCCCCCTCCACGGTCTGATGACTGTTGGTTTCGATGTCTGCCATTCACCGAAGAAT 123

Qy       1947 CGGAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTAC 2006
                ||| | | || || || ||  |||| | ||||| |   || |    | |  |||
Db         124 AAGAACAAGTCATATGGGGCTTTCGTTGCAACCATGGACCAGAAGGAGTCTTTCCGCTAC 183

Qy       2007 TTCAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATG 2066
              |||  ||| || |  ||  ||          | | | |  |  |     | | |  ||
Db         184 TTCTCCACCGTAAACGAACACATTAAGGGCCAAGAGTTGTCCGAGCAGATGTCGGTGAAC 243

Qy       2067 ATAGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGTA 2126
              || ||    |||||| |    |||||  |  | ||  |      |||||   || ||  |
Db         244 ATGGCGTGCGCCCTGAGGTCGTATCAGGAGCAACACCGTTCTTTGCCAGAGCGCATTCTC 303

Qy       2127 TTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAAG 2186
              || | ||| ||||||||  ||   ||    ||  | ||| |  |  |    || || ||
Db         304 TTCTTTCGCGACGGTGTTGGCGATGGTCAGCTCTACCAGGTGGTAAACAGCGAGGTGAAC 363

Qy       2187 GACATCATTGAGAAGTTG---AAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAA 2243
```

```
Db        364 ACCCTAAAGGACAGGCTCGACGAAATTTACAAATCAGCTGGCAAACAGGAGGGCTGTCGC 423

Qy       2244 TTAGCTTATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAAT 2303

Db        424 ATGACATTTATAATTGTATCCAAGCGCATTAATTCTCGCTACTTTACTGGGCATCGCAAC 483

Qy       2304 CCTCCGCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTAT 2363

Db        484 CCAGTTCCGGGCACTGTAGTCGATGACGTTATTACCTTGCCAGAGCGCTACGACTTCTTC 543

Qy       2364 CTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTAT 2423

Db        544 CTAGTGTCCCAGGCTGTTCGGATAGGAACTGTGTCGCCTACCAGCTACAATGTTATTTCT 603

Qy       2424 AGCAGCATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTG 2483

Db        604 GACAACATGGGACTAAACGCCGATAAGCTGCAGATGCTCTCCTATAAGATGACCCATATG 663

Qy       2484 TACTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTA 2543

Db        664 TACTACAATTACAGCGGAACCATACGAGTCCCCGCTGTCTGCCATTACGCCCACAAATTG 723

Qy       2544 G 2544

Db        724 G 724
```

Search completed: July 21, 2003, 22:52:59
Job time : 4006.44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2003, 19:01:27 ; Search time 29.5674 Seconds
(without alignments)
3799.123 Million cell updates/sec

Title: US-09-873-737A-2
Perfect score: 4389
Sequence: 1 MADDQGRGRRRPXNEDDSST.....VGTNLHSIPQNALEKKFYYL 843

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4383	99.9	843	21	AAY90233	Drosophila piwi pr
2	4361	99.4	843	22	ABB62084	Drosophila melanog
3	1801.5	41.0	866	22	ABB62102	Drosophila melanog
4	1406	32.0	861	21	AAY90235	Human piwi protein
5	1385.5	31.6	862	21	AAY90234	Mouse piwi protein
6	875.5	19.9	580	22	AAU07866	Polypeptide sequen
7	788	18.0	530	22	AAB94209	Human protein sequ
8	582	13.3	371	22	AAB92722	Human protein sequ
9	458.5	10.4	950	22	ABB67046	Drosophila melanog
10	458.5	10.4	984	22	ABB62573	Drosophila melanog
11	458.5	10.4	984	22	ABB67045	Drosophila melanog
12	448	10.2	860	22	AAM80292	Human protein SEQ
13	448	10.2	860	22	AAB94291	Human protein sequ
14	436.5	9.9	906	21	AAG51563	Arabidopsis thalia
15	436.5	9.9	983	21	AAG51562	Arabidopsis thalia
16	436.5	9.9	988	21	AAG51561	Arabidopsis thalia
17	432.5	9.9	782	22	AAB93139	Human protein sequ
18	431	9.8	1048	21	AAG42118	Arabidopsis thalia
19	423	9.6	1064	21	AAG52593	Arabidopsis thalia
20	419	9.5	1145	22	AAE09770	Drosophila melanog
21	410.5	9.4	962	22	ABB63278	Drosophila melanog
22	395.5	9.0	868	21	AAG42119	Arabidopsis thalia
23	387.5	8.8	884	21	AAG52594	Arabidopsis thalia
24	370	8.4	341	22	ABG05957	Novel human diagno
25	367.5	8.4	820	21	AAG42120	Arabidopsis thalia
26	359.5	8.2	836	21	AAG52595	Arabidopsis thalia
27	315	7.2	938	22	AAB82553	Neurospora crassa
28	254.5	5.8	252	22	AAB97112	Translation initia
29	253	5.8	406	21	AAG54509	Zea mays protein f
30	241	5.5	199	22	ABG05956	Novel human diagno
31	231.5	5.3	352	21	AAG54511	Zea mays protein f
32	231.5	5.3	353	21	AAG54510	Zea mays protein f
33	229.5	5.2	326	21	AAG52610	Arabidopsis thalia
34	223	5.1	140	22	ABG12914	Novel human diagno
35	213.5	4.9	227	21	AAG52611	Arabidopsis thalia
36	205	4.7	1020	22	AAU01856	C. elegans RNA int
37	187.5	4.3	108	22	AAO04945	Human polypeptide
38	182	4.1	232	22	AAM85465	Human immune/haema
39	171	3.9	747	21	AAG25621	Arabidopsis thalia
40	171	3.9	749	21	AAG25620	Arabidopsis thalia
41	168	3.8	130	21	AAB58977	Breast and ovarian
42	143.5	3.3	118	22	AAU32904	Novel human secret
43	135.5	3.1	567	21	AAG25622	Arabidopsis thalia
44	132	3.0	71	23	ABG40429	Human peptide enco
45	124.5	2.8	5147	22	ABB59831	Drosophila melanog

ALIGNMENTS

RESULT 1
AAY90233
ID AAY90233 standard; Protein; 843 AA.
XX
AC AAY90233;
XX
DT 29-AUG-2000 (first entry)
XX
DE Drosophila piwi protein.
XX
KW Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW anaemia; immunodeficiency; male infertility; Drosophila.
XX
OS Drosophila sp.
XX
FH Key Location/Qualifiers
FT Misc-difference 13
FT /label= Leu, Ile
FT /note= "encoded by NTT"
FT Misc-difference 106
FT /label= Leu, Ile
FT /note= "encoded by NTA"
FT Misc-difference 785
FT /label= Leu, Ile
FT /note= "encoded by NTC"
XX
PN WO200032039-A1.
XX
PD 08-JUN-2000.
XX
PF 03-DEC-1999; 99WO-US28764.
XX
PR 04-DEC-1998; 98US-0110901.

XX
PA (UYDU-) UNIV DUKE.
XX
PI Lin H;
XX
DR WPI; 2000-412085/35.
DR N-PSDB; AAA07586.
XX
PT Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT therapy of diseases such as cancer and in various research and
PT diagnostic applications -
XX
PS Claim 4; Page 171-175; 201pp; English.
XX
CC This sequence represents the Drosophila piwi family protein, designated
CC piwi. The piwi family nucleic acids and polypeptides are used in gene
CC therapy of diseases such as cancer and also in various research and
CC diagnostic applications. The sequences can also be used to treat
CC tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ Sequence 843 AA;

Query Match 99.9%; Score 4383; DB 21; Length 843;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MADDQGRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MADDQGRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAP 60

Qy         61 RREGGPPERKPWGDQYDYLNTRPVELVSKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RREGGPPERKPWGDQYDYLNTRPVELVSKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVE 120

Qy        121 FVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 FVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISI 180

Qy        181 KFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETS 240

Qy        241 IRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTY 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTY 300

Qy        301 RINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASEL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASEL 360

Qy        361 VVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLR 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLR 420

Qy        421 DWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDGLDRWAV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDGLDRWAV 480

Qy        481 IAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCVRSDPKLIL 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 IAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCVRSDPKLIL 540

Qy        541 CLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLGYTPWM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLGYTPWM 600

Qy        601 IELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWP 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWP 660

Qy        661 MIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLSPPQ 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 MIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLSPPQ 720

Qy        721 LAYIVVTRSMNTRFFLNGQNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLY 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 LAYIVVTRSMNTRFFLNGQNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLY 780

Qy        781 SSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 SSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKF 840

Qy        841 YYL 843
              |||
Db        841 YYL 843
```

RESULT 2
ABB62084
ID ABB62084 standard; Protein; 843 AA.
XX
AC ABB62084;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 13044.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL06187.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 13044; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 843 AA;

Query Match 99.4%; Score 4361; DB 22; Length 843;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 836; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

```
Qy          1 MADDQGRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAP 60
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MADDQGRGRRRPLNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAP 60
```

```
Qy         61 RREGGPPERKPWGDQYDYLNTRPVELVSKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVE 120
Db         61 RREGGPTERKPWGDQYDYLNTRPAELVSKKGTDGVPVMLQTNFFRLKTKPEWRIVHYHVE 120

Qy        121 FVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISI 180
Db        121 FEPSIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISI 180

Qy        181 KFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETS 240
Db        181 KFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETS 240

Qy        241 IRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTY 300
Db        241 IRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTY 300

Qy        301 RINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASEL 360
Db        301 RINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASEL 360

Qy        361 VVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLR 420
Db        361 VVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLR 420

Qy        421 DWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDGLDRWAV 480
Db        421 DWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDGLDRWAV 480

Qy        481 IAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCVRSDPKLIL 540
Db        481 IAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCVRSDPKLIL 540

Qy        541 CLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLGYTPWM 600
Db        541 CLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLGYTPWM 600

Qy        601 IELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWP 660
Db        601 IELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWP 660

Qy        661 MIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLSPPQ 720
Db        661 MIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLSPPQ 720

Qy        721 LAYIVVTRSMNTRFFLNGQNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLY 780
Db        721 LAYIVVTRSMNTRFFLNGQNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLY 780

Qy        781 SSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKF 840
Db        781 SSMGLSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKF 840

Qy        841 YYL 843
Db        841 YYL 843
```

RESULT 3
ABB62102
ID ABB62102 standard; Protein; 866 AA.
XX
AC ABB62102;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 13098.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL06205.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 13098; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 866 AA;

Query Match 41.0%; Score 1801.5; DB 22; Length 866;
Best Local Similarity 42.4%; Pred. No. 1.4e-158;
Matches 374; Conservative 159; Mismatches 279; Indels 71; Gaps 13;

```
Qy          5 QGRGR-RRPXN------------EDDSSTSRGS---------GD---GPRVK------V 32
Db         11 RGRGRGRKPNNVEANRGFAPSLGQKSDPSHSEGNQASGGNGGGGDAQVGPSIEKSSLSAV 70

Qy         33 FRGSSSGDPRADPRIEASRERRALEEAPRREGGPPERKPWGDQYDYLNTRPVELVSKKGT 92
Db         71 QMHKSEGDPRG-----SVRGRRLIT------------------DLVYSRPPGMTSKKGV 106

Qy         93 DGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMGVLSNHANLLGSGYLFDGLQ 152
Db        107 VGTHITVQANYFKVLKRPNWTIYQYRVDFTPDVEATRLRRSFLYEHKGILG-GYIFDGTN 165

Qy        153 LFTTRKFE-----QEITVLSGKSKLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKG 207
Db        166 MFCINQFKAVQDSPYVLELVTKSRAGENIEIKIKAVGSVQSTDAEQFQVLNLILRRAMEG 225

Qy        208 LNLELVGRNLFDPRAKIEIREFKMELWPGYETSIRQHEKDILLGTEITHKVMRTETIYDI 267
Db        226 LDLKLVSRYYYDPQAKINLENFRMQLWPGYQTSIRQHENDILLCSEICHKVMRTETLYNI 285

Qy        268 MRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTFSCKGRDISFVEY 327
Db        286 LSDAIRDSDDYQSTFKRAVMGMVILTDYNNKTYRIDDVDFQSTPLCKFKTNDGEISYVDY 345

Qy        328 YLTKYNIRIRDHNQPLLISKNRDKALKTNASELVVLIPELCRVTGLNAEMRSNFQLMRAM 387
Db        346 YKKRYNIIIRDLKQPLVMSRPTDKNIRGGNDQAIMIIPELARATGMTDAMRADFRTLRAM 405

Qy        388 SSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHN 447
Db        406 SEHTRLNPDRRIERLRMFNKRLKSCKQSVETLKSWNIELDSALVEIPARVLPPEKILFGN 465

Qy        448 GKV-PAGENADWQRHFRDQRMLTTPSDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGL 506
```

```
Db        466 QKIFVCDARADWTNEFRTCSMFKNVH--INRWYVITPSRNLRETQEFVQMCIRTASSMKM 523

Qy        507 RIRSPQEFIIYDDRTGTYVRAMDDCVRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPT 566

Db        524 NICNPIYEEIPDDRNGTYSQAIDNAAANDPQIVMVVMRSPNEEKYSCIKKRTCVDRPVPS 583

Qy        567 QVVTLKT-----TKNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRD 621

Db        584 QVVTLKVIAPRQQKPTGLMSIATKVVIQMNAKLMGAPWQVVIPLHGLMTVGFDVCHSPKN 643

Qy        622 RKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIV 681

Db        644 KNKSYGAFVATMDQKESFRYFSTVNEHIKGQELSEQMSVNMACALRSYQEQHRSLPERIL 703

Qy        682 FYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARV-QLSPPQLAYIVVTRSMNTRFFLNGQN 740

Db        704 FFRDGVGDGQLYQVVNSEVNTLKDRLDEIYKSAGKQEGCRMTFIIVSKRINSRYFTGHRN 763

Qy        741 PPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHL 800

Db        764 PVPGTVVDDVITLPERYDFFLVSQAVRIGTVSPTSYNVISDNMGLNADKLQMLSYKMTHM 823

Qy        801 YYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKFYYL 843

Db        824 YYNYSGTIRVPAVCHYAHKLAFLVAESINRAPSAGLQNQLYFL 866
```

RESULT 4
AAY90235
ID AAY90235 standard; Protein; 861 AA.
XX
AC AAY90235;
XX
DT 29-AUG-2000 (first entry)
XX
DE Human piwi protein, designated hiwi.
XX
KW Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW anaemia; immunodeficiency; male infertility; human.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 76
FT /label= Leu, Ile
FT /note= "encoded by NTA"
FT Misc-difference 303
FT /label= Leu, Ile
FT /note= "encoded by NTA"
FT Misc-difference 735
FT /label= Leu, Ile
FT /note= "encoded by NTA"
XX
PN WO200032039-A1.
XX
PD 08-JUN-2000.
XX
PF 03-DEC-1999; 99WO-US28764.
XX
PR 04-DEC-1998; 98US-0110901.
XX
PA (UYDU-) UNIV DUKE.
XX
PI Lin H;
XX
DR WPI; 2000-412085/35.
DR N-PSDB; AAA07588.
XX
PT Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT therapy of diseases such as cancer and in various research and
PT diagnostic applications -
XX
PS Claim 4; Page 189-194; 201pp; English.

XX
CC This sequence represents the human piwi family protein, designated
CC hiwi. The piwi family nucleic acids and polypeptides are used in gene
CC therapy of diseases such as cancer and also in various research and
CC diagnostic applications. The sequences can also be used to treat
CC tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ Sequence 861 AA;

Query Match 32.0%; Score 1406; DB 21; Length 861;
Best Local Similarity 35.8%; Pred. No. 1.2e-121;
Matches 308; Conservative 172; Mismatches 312; Indels 68; Gaps 20;

```
Qy          6 GRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG 65

Db         48 GRGRQR-------GTAGGTAKSQGLQISAG-------------------FQELSXAERG 80

Qy         66 PPERKPWGDQYDY-LNTRP-VELV--SKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEF 121

Db         81 GRRR----DFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDY 136

Qy        122 VPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIK 181

Db        137 NPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITIT 196

Qy        182 FVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSI 241

Db        197 LTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSI 256

Qy        242 RQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRT 299

Db        257 LQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKT 316

Qy        300 YRINDVDEGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNA 357

Db        317 YRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTL 376

Qy        358 SELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPE 414

Db        377 PGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQ- 435

Qy        415 SVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTP 471

Db        436 --RELRDWGLSFDSNLLSFSGRILQTEKI--HQGGKTFDYNPQFADWSKETRGAPLISVK 491

Qy        472 SDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDC 531

Db        492 P--LDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRK-AIMIEVDDRTEAYLRVLQQK 548

Qy        532 VRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQL 590

Db        549 VTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQM 608

Qy        591 NCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSA 650

Db        609 NCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASIN-EGMTRWFSRCIFQDR 667

Qy        651 FDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKT- 709

Db        668 GQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSI 727

Qy        710 --EYARVQLSPPQLAYIVVTRSMNTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVS 763

Db        728 GRGY------NPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVS 781

Qy        764 QQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATL 823

Db        782 QAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFL 841

Qy        824 VGTNLHSIPQNALEKKFYYL 843

Db        842 VGQSIHREPNLSLSNRLYYL 861
```

```
RESULT 5
AAY90234
ID   AAY90234 standard; Protein; 862 AA.
XX
AC   AAY90234;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Mouse piwi protein, designated miwi.
XX
KW   Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW   anaemia; immunodeficiency; male infertility; mouse; ds.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 90
FT                   /label= Leu, Ile
FT                   /note= "encoded by NTT"
FT   Misc-difference 216
FT                   /note= "unspecified amino acid; encoded by NTC"
FT   Misc-difference 383
FT                   /label= Leu, Ile
FT                   /note= "encoded by NTC"
FT   Misc-difference 816
FT                   /label= Leu, Ile
FT                   /note= "encoded by NTC"
XX
PN   WO200032039-A1.
XX
PD   08-JUN-2000.
XX
PF   03-DEC-1999;   99WO-US28764.
XX
PR   04-DEC-1998;   98US-0110901.
XX
PA   (UYDU-) UNIV DUKE.
XX
PI   Lin H;
XX
DR   WPI; 2000-412085/35.
DR   N-PSDB; AAA07587.
XX
PT   Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT   therapy of diseases such as cancer and in various research and
PT   diagnostic applications -
XX
PS   Claim 4; Page 180-185; 201pp; English.
XX
CC   This sequence represents the mouse piwi family protein, designated
CC   miwi. The piwi family nucleic acids and polypeptides are used in gene
CC   therapy of diseases such as cancer and also in various research and
CC   diagnostic applications. The sequences can also be used to treat
CC   tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ   Sequence   862 AA;

  Query Match             31.6%;  Score 1385.5;  DB 21;  Length 862;
  Best Local Similarity   35.7%;  Pred. No. 9.9e-120;
  Matches  306;  Conservative  173;  Mismatches  318;  Indels   59;  Gaps   19;

Qy          6 GRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG 65
              ||||:|      :| |:                    : :| |  :  :| |'     ||
Db         48 GRGRQRGMVVGATSKSQ--------------------ELQISAGFQELSLAE----RGG 82

Qy         66 PPERKPWGDQYDYLNTRP-VELV--SKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFV 122
                |: : |    :|||  :: |  || |: |: | | || ||  ::|:| :  ||:::
Db         83 --RRRDFHD--XGVNTRQNLDHVKESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYN 138

Qy        123 PTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIKF 182
               | :|  |:|  :|  | :|:|  : |||  ||  :: : ::| :  :::     :|:|
Db        139 PLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVTEVFSQTRNGEHVRITITL 198

Qy        183 VGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSIR 242
               :   |  ||  |:  || :| :||: :||| ::|   |:|   :: :|||: |||
Db        199 TNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSIL 258

Qy        243 QHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRTY 300
              |:| :|:| |:::|||:|:||: | |          |  |::|   :: ||||| |||:||
Db        259 QYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTY 318

Qy        301 RINDVDFGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNAS 358
              |::|:|: | |||||    | ::||:||| :|| | | ||:|:|:  | :
Db        319 RVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLP 378

Qy        359 ELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPES 415
               : ||||| :||| :||::| :|: :: :||: |:||     ||  : |:  |
Db        379 GPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQ-- 436

Qy        416 VKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTPS 472
               : |||| :  | |:    |||:  : |  | |       |   ||| :  |   :::
Db        437 -RELRDWGLSFDSNLLSFSGRILQSEKI--HQGGKTFDYNPQFADWSKETRGAPLISVKP 493

Qy        473 DGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCV 532
              || | :|  :||     :|: :|::    ||:::     |  |||| |:||:   |
Db        494 --LDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKK-AIMIEVDDRTEAYLRALQQKV 550

Qy        533 RSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQLN 591
              || ::::||: ::  ::| :|||    |   |:| |  :|  | :::|:||||||:|:|
Db        551 TSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMN 610

Qy        592 CKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAF 651
              ||:|   | ::: |   | :| |    |   :|:    :||:: :  : :||
Db        611 CKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASIN-EGMTRWFSRCVFQDRG 669

Qy        652 DVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEY 711
              | : |   :  ||| :   :  :|||:: |||||  | || |  :||   :: ||:
Db        670 QELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVG 729

Qy        712 ARVQLSPPQLAYIVVTRSMNTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVSQQVR 767
                    |:|  ||| : :| |||       ||| |||::|  :| || |||::||| ||
Db        730 RGYN---PRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVR 786

Qy        768 QGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTN 827
              |:|||| |||:| | |  |: :|:|||| ||:|||| |  |||| |||| ||| ||| :
Db        787 SGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVGQS 846

Qy        828 LHSIPQNALEKKFYYL 843
              :|  |  :|  : |||
Db        847 IHREPNLSLSNRLYYL 862


RESULT 6
AAU07866
ID   AAU07866 standard; Protein; 580 AA.
XX
AC   AAU07866;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Polypeptide sequence for mammalian Spg16.
XX
KW   Mammalian; reproductive-specific protein; male infertility;
KW   spermatogenesis; sperm count disorder; anti infertility; reproduction.
XX
OS   Mammalia.
XX
PN   WO200166752-A2.
XX
PD   13-SEP-2001.
XX
PF   07-MAR-2001; 2001WO-US07371.
XX
PR   07-MAR-2000; 2000US-0187518.
PR   12-JAN-2001; 2001US-0261557.
```

XX
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
XX
PI Wang PJ, Page DC;
XX
DR WPI; 2001-570774/64.
DR N-PSDB; AAS13630.
XX
PT Novel reproduction-specific protein, useful for treating disorders of
PT reduced sperm count, enhancing/increasing sperm count and/or sperm
PT activity -
XX
PS Claim 22; Fig 16; 151pp; English.
XX
CC The present invention relates to the isolation of novel mammalian and
CC human reproductive-specific proteins (AAU07859-AAU07899), and the
CC nucleic acids encoding them. The nucleic acids encoding
CC reproductive-specific proteins are useful for diagnosing infertility
CC which is a result of reduced sperm count, reduced sperm motility,
CC malformed sperm or combinations of these. The sequences of the invention
CC are useful as markers for spermatogonial cells, for identifying genes or
CC proteins characteristic of male infertility, diagnosing or aiding in
CC the diagnosis of infertility in men, and for contraception in which
CC sperm production or sperm count is reduced or defective sperm is
CC produced. Antibodies to reproductive-specific proteins are useful for
CC determining the presence of these proteins in a sample obtained from a
CC man being assessed for infertility, for identifying the expression of
CC genes in particular cell type or particular developmental stage, for
CC studies of spermatogenesis, and for immunofluorescence of germ cells or
CC in Western blots for assessing the presence of the protein the antibody
CC binds. The sequences of the invention are also useful for treating
CC disorders of reduced sperm count, and for increasing sperm count and/or
CC sperm activity. The nucleic acids of the invention are useful in gene
CC therapy. AAU07859-AAU07882 represent the mammalian reproduction-specific
CC proteins of the present invention.
XX
SQ Sequence 580 AA;

Query Match 19.9%; Score 875.5; DB 22; Length 580;
Best Local Similarity 33.8%; Pred. No. 2.2e-72;
Matches 197; Conservative 120; Mismatches 243; Indels 23; Gaps 10;

```
Qy        274 NPARHQDEVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTF-SCKGRDISFVEYYLTKY 332
              |    |||    ::  ||:| ||||||||:|||: :||| :|    |::|:|:|||   |
Db          8 NKEHFQDECSKLLVGSIVITRYNNRTYRIDDVDWNKTPKDSFVMSDGKEITFLEYYSKNY 67

Qy        333 NIRIRDHNQPLLISKNRDKALKTNASEL----VVLIPELCRVTGLNAEMRSNFQLMRAMS 388
              | ::: :|||||  :|    : |   |    ::|:|||  :||:  :|: :|: |: ::
Db         68 GITVKEDDQPLLI--HRPSERQNNHGMLLKGEILLLPELSFMTGIPEKMKKDFRAMKDLT 125

Qy        389 SYTRMNPKQRTDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNG 448
              ::|||    |     |:     :   |  | : | |:| :::||::  : |   |
Db        126 QQINLSPKQHHGALECLLQRISQNETASNELTRWGLSLHKDVHKIEGRLLPMERINLRNT 185

Qy        449 KVPAGENADWQRHF-RDQRMLTTPSDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLR 507
                |  :| :   ||  :|| |   :  ||:  |:|   : | |:: | : |  :|:|
Db        186 SFVTSEGLNWVKEVTRDASILTIP---MHFWALFYPKRAMDQARELVNMLEKIAGPIGMR 242

Qy        508 IRSPQEFIIYDDRTGTYVRAMDDCVRSDPK--LILCLVPNDNAERYSSIKKRGYVDRAVP 565
                |    : |||  ||:| :   :  : |  :::|::     : | :|||   |   ||
Db        243 TSPPAWVELKDDRIETYIRTIQSLLGVEGKIQMVVCIIMGTRDDLYGAIKKLCCVQSPVP 302

Qy        566 TQVVTLKTT-KNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKR 624
              :||: ::|  :   | |:| || :|:|||||  `| :::||  || || |:         |
Db        303 SQVINVRTIGQPTRLRSVAQKILLQMNCKLGGELWGVDIPLKQLMVIGMDVYHDPSRGMR 362

Qy        625 AYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYR 684
              :    :||::|     |   | :    ::: ::|   :  :|::|  :  || :|| ||
Db        363 SVVGFVASINLTLTKWYSRVVFQMPHQEIV-DSLKLCLVGSLKKYYEVNHCLPEKIVVYR 421

Qy        685 DGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLSPPQLAYIVVTRSMNTRFFLNGQN---- 740
              |||| | || :  :|    | :|:  :       |::   || : ::|  :|   :
Db        422 DGVSDGQLKTVANYE----IPQLQKCFEAFDNYHPKMVVFVVQKKISTNLYLAAPDHFVT 477

Qy        741 PPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHL 800
              | |||:||  ||  |  ||||::  ||||   || |  : ::   ||: ||:||:|:||:
Db        478 PSPGTVVDHTITSCEWVDFYLLAHHVRQGCGIPTHYICVLNTANLSPDHMQRLTFKLCHM 537

Qy        801 YYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKFYYL 843
              |:|| || |||| |:|| ||| | |  ||  |   |    ::|
Db        538 YWNWPGTIRVPAPCKYAHKLAFLSGQILHHEPAIQLCGNLFFL 580
```

RESULT 7
AAB94209
ID AAB94209 standard; Protein; 530 AA.
XX
AC AAB94209;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:14559.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 14559; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX

SQ Sequence 530 AA;

Query Match 18.0%; Score 788; DB 22; Length 530;
Best Local Similarity 32.7%; Pred. No. 2.9e-64;
Matches 176; Conservative 115; Mismatches 222; Indels 26; Gaps 10;

```
Qy        319 GRDISFVEYYLTKYNIRIRDHNQPLLISK------NRDKALKTNASELVVLIPELCRVTG 372
              |::|:|:|||   | | ::: :|||||  :       |    ||      ::|:|||  :||
Db          4 GKEITFLEYYSKNYGITVKEEDQPLLIHRPSERQDNHGMLLKGE----ILLLPELSFMTG 59

Qy        373 LNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTE 432
              :  :|: :|: |: ::    ::|||    |      |:      :   |  | : | |:| :
Db         60 IPEKMKKDFRAMKDLAQQINLSPKQHHSALECLLQRIAKNEAATNELMRWGLRLQKDVHK 119

Qy        433 VQGRIIGQQNIVFHNGKVPAGENADWQRHF-RDQRMLTTPSDGLDRWAVIAPQRNSHELR 491
              ::||::  : |   |       :  :| :   ||  :|| |   :  ||:  |:|   : |
Db        120 IEGRVLPMERINLKNTSFITSQELNWVKEVTRDPSILTIP---MHFWALFYPKRAMDQAR 176

Qy        492 TLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCVRSDPK--LILCLVPNDNAE 549
              |:: | : |  :|:|:  |    : |||  |||| :   : :: |  :::|::      :
Db        177 ELVNMLEKIAGPIGMRMSPPAWVELKDDRIETYVRTIQSTLGAEGKIQMVVCIIMGPRDD 236

Qy        550 RYSSIKKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGL 608
              | :|||   |   ||:||| ::|  :   | |:| || :|:|||||   | :::||  |
Db        237 LYGAIKKLCCVQSPVPSQVVNVRTIGQPTRLRSVAQKILLQINCKLGGELWGVDIPLKQL 296

Qy        609 MTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQ 668
              | || |:        |:    :||::|     |   | :    ::: ::|   :  :|::
Db        297 MVIGMDVYHDPSRGMRSVVGFVASINLTLTKWYSRVVFQMPHQEIV-DSLKLCLVGSLKK 355

Qy        669 YQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLSPPQLAYIVVTR 728
              :   :  || :|| |||||| | || :  :|:   ::|    :   |   |::   || :
Db        356 FYEVNHCLPEKIVVYRDGVSDGQLKTVANYEIPQ-LQKCFEAFENYQ---PKMVVFVVQK 411

Qy        729 SMNTRFFL----NGQNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMG 784
              ::|  :|     |   | |||:||  ||  |  ||||::  ||||   || |  : ::
Db        412 KISTNLYLAAPQNFVTPTPGTVVDHTITSCEWVDFYLLAHHVRQGCGIPTHYVCVLNTAN 471

Qy        785 XSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKFYYL 843
              ||: ||:||:|:||:|:|| || |||| |:|| ||| | |  ||  |   | :  ::|
Db        472 LSPDHMQRLTFKLCHMYWNWPGTIRVPAPCKYAHKLAFLSGHILHHGPAIQLCENLFFL 530
```

RESULT 8
AAB92722
ID AAB92722 standard; Protein; 371 AA.
XX
AC AAB92722;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:11144.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 11144; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 371 AA;

Query Match 13.3%; Score 582; DB 22; Length 371;
Best Local Similarity 34.9%; Pred. No. 2.8e-45;
Matches 130; Conservative 77; Mismatches 154; Indels 12; Gaps 6;

```
Qy        478 WAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCVRSDPK 537
              ||:  |:|   : | |:: | : |  :|:|:  |    : |||  |||| :   : :: |
Db          4 WALFYPKRAMDQARELVNMLEKIAGPIGMRMSPPAWVELKDDRIETYVRTIQSTLGAEGK 63

Qy        538 --LILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQLNCKL 594
                :::|::     : | :| |   |   ||:||| ::|  :   | |:| || :|:||||
Db         64 IQMVVCIIMGPRDDLYGAINKLCCVQSPVPSQVVNVRTIGQPTRLRSVAQKILLQINCKL 123

Qy        595 GYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVL 654
              |   | :::||  || || |:        |:    :||::|     |   | :    :::
Db        124 GGELWGVDIPLKQLMVIGMDVYHDPSRGMRSVVGFVASINLTLTKWYSRVVFQMPHQEIV 183

Qy        655 ANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARV 714
              ::|   :  :|:::   :  || :|| |||||| | || :  :|:   ::|    :
Db        184 -DSLKLCLVGSLKKFYEVNHCLPEKIVVYRDGVSDGQLKTVANYEIPQ-LQKCFEAFENY 241

Qy        715 QLSPPQLAYIVVTRSMNTRFFL----NGQNPPPGTIVDDVITLPERYDFYLVSQQVRQGT 770
              |   |::   || : ::|  :|     |   | |||:||  ||  |  ||||::  ||||
Db        242 Q---PKMVVFVVQKKISTNLYLAAPQNFVTPTPGTVVDHTITSCEWVDFYLLAHHVRQGC 298

Qy        771 VSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHS 830
              || |  : ::   ||: ||:||:|:||:|:|| || |||| |:|| ||| | |  ||
Db        299 GIPTHYVCVLNTANLSPDHMQRLTFKLCHMYWNWPGTIRVPAPCKYAHKLAFLSGHILHH 358

Qy        831 IPQNALEKKFYYL 843
              |   | :  ::|
Db        359 EPAIQLCENLFFL 371
```

RESULT 9
ABB67046
ID ABB67046 standard; Protein; 950 AA.
XX
AC ABB67046;

```
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 27930.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL11149.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 27930; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   950 AA;

  Query Match             10.4%;  Score 458.5;  DB 22;  Length 950;
  Best Local Similarity   23.9%;  Pred. No. 4.6e-33;
  Matches  200;  Conservative  144;  Mismatches  343;  Indels  151;  Gaps   39;

Qy         91 GTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMGVLSN--HANLLGSGYL- 147
              | :| |::|: | |:  | |   : || :   |     :|   ::    ||     | |
Db        108 GREGRPIVLRANHFQV-TMPRGYVHHYDINIQPDKCPRKVNREIIETMVHAYSKIFGVLK 166

Qy        148 --FDGLQLFTTR--------KFEQEITVLSGKSKLDIEYKISIKFVGFIS------CAEP 191
                |||     ||        : | |:| | |: | |  ::::||:   :|        |
Db        167 PVFDGRNNLYTRDPLPIGNERLELEVT-LPGEGK-DRIFRVTIKWQAQVSLFNLEEALEG 224

Qy        192 RFLQV-------LNLILRRSMKGLNLELVGRNLF-DPRAKIEIREFKMELWPGYETSIRQ 243
              |  |:       |::::|  :  :    |||: |  |             |:| |:  |:|
Db        225 RTRQIPYDAILALDVVMRH-LPSMTYTPVGRSFFSSPEGYYHPLGGGREVWFGFHQSVRP 283

Qy        244 HEKDILLGTEITHKVM-RTETIYDIMRRCSHNPARHQDEVRVNVLDL------------ 289
              :  ::|  :::     : : : | |  |     |  :| |  : |
Db        284 SQWKMMLNIDVSATAFYKAQPVIDFM--CEVLDIRDINEQRKPLTDSQRVKFTKEIKGLK 341

Qy        290 --IVLTDYNNRTYRINDV-------------DFGQTPKSTFSCKGRDISFVEYYLTKYN 333
                |       | ||: :|              : ||| : | :          :|:| ||
Db        342 IEITHCGQMRRKYRVCNVTRRPAQMQSFPLQLENGQTVECTVA---------KYFLDKYR 392

Qy        334 IRIRDHNQPLLISKNRDKALKTNASELVVLIP-ELCRVTGLNAEMRSNFQLMRAMSSYTR 392
              :::|  : |          |:         :| |:| :    |  |   : :  | : |
Db        393 MKLRYPHLP---------CLQVGQEHKHTYLPLEVCNIV---AGQRC-IKKLTDMQTSTM 439

Qy        393 MNPKQRT--DRLRAFNHRLQNTP-ESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGK 449
              :   |:  || |  |: ::      :   :::: : :  :: ||:||::    : :  |:
Db        440 IKATARSAPDREREINNLVKRADFNNDSYVQEFGLTISNSMMEVRGRVLPPPKLQY-GGR 498

Qy        450 VPAGENADWQRHFRDQR--MLTTPSDG-------------LDRWAV--IAPQRNSHE-- 489
              |  |     |: |  |    | :|: |               :  ||:   ||||   |
Db        499 VSTGLTG--QQLFPPQNKVSLASPNQGVWDMRGKQFFTGVEIRIWAIACFAPQRTVREDA 556

Qy        490 LRTLLDSLYRAASGMGLRIRSPQEFIIY---DDRTGTYVRAMDDCVRSDPKLILCLVPND 546
              ||     | : ::  |: |     |  |    |:     | :          :|:: ::|
Db        557 LRNFTQQLQKISNDAGMPIIGQPCFCKYATGPDQVEPMFRYL-KITFPGLQLVVVVLPG- 614

Qy        547 NAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLG---------YT 597
                |: :|: |     : || |  |    |::     : : :::| |||
Db        615 KTPVYAEVKRVGDTVLGMATQCVQAKNV-NKTSPQTLSNLCLKINVKLGGINSILVPSIR 673

Qy        598 PWMIELPLSGLMTIGFDIA--KSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLA 655
              | :   |   :: :| |:    :  ::| :  |:: |||    | | :||        :
Db        674 PKVFNEP---VIFLGADVTHPPAGDNKKPSIAAVVGSMDAHP-SRYAATVRVQQHRQEII 729

Qy        656 NTLWPMIAKALRQ-YQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIE---KLKTEY 711
                |  |: : |   |:       | ||: |||||| |    : : |:  | |   ||: ||
Db        730 QELSSMVRELLIMFYKSTGGYKPHRIILYRDGVSEGQFPHVLQHELTAIREACIKLEPEY 789

Qy        712 ARVQLSPPQLAYIVVTRSMNTRFFL--------NGQNPPPGTIVDDVITLPERYDFYLVS 763
                  | : :|||  :  :|| |            | | || ||  || |  :|||| |
Db        790 R------PGITFIVVQKRHHTRLFCAEKKEQSGKSGNIPAGTTVDVGITHPTEFDFYLCS 843

Qy        764 QQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLA 821
              |  |||  |: |:||:       :::| ||||::|| |   : :  :||   ||  :|
Db        844 HQGIQGTSRPSHYHVLWDDNHFDSDELQCLTYQLCHTYVRCTRSVSIPAPAYYAHLVA 901


RESULT 10
ABB62573
ID   ABB62573 standard; Protein; 984 AA.
XX
AC   ABB62573;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 14511.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL06676.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 14511; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
```

CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 984 AA;

Query Match 10.4%; Score 458.5; DB 22; Length 984;
Best Local Similarity 23.9%; Pred. No. 4.9e-33;
Matches 200; Conservative 144; Mismatches 343; Indels 151; Gaps 39;

```
Qy         91 GTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMGVLSN--HANLLGSGYL- 147
Db        142 GREGRPIVLRANHFQV-TMPRGYVHHYDINIQPDKCPRKVNREIIETMVHAYSKIFGVLK 200

Qy        148 --FDGLQLFTTR--------KFEQEITVLSGKSKLDIEYKISIKFVGFIS------CAEP 191
Db        201 PVFDGRNNLYTRDPLPIGNERLELEVT-LPGEGK-DRIFRVTIKWQAQVSLFNLEEALEG 258

Qy        192 RFLQV-------LNLILRRSMKGLNLELVGRNLF-DPRAKIEIREFKMELWPGYETSIRQ 243
Db        259 RTRQIPYDAILALDVVMRH-LPSMTYTPVGRSFFSSPEGYYHPLGGGREVWFGFHQSVRP 317

Qy        244 HEKDILLGTEITHKVM-RTETIYDIMRRCSHNPARHQDEVRVNVLDL------------- 289
Db        318 SQWKMMLNIDVSATAFYKAQPVIDFM--CEVLDIRDINEQRKPLTDSQRVKFTKEIKGLK 375

Qy        290 --IVLTDYNNRTYRINDV--------------DFGQTPKSTFSCKGRDISFVEYYLTKYN 333
Db        376 IEITHCGQMRRKYRVCNVTRRPAQMQSFPLQLENGQTVECTVA---------KYFLDKYR 426

Qy        334 IRIRDHNQPLLISKNRDKALKTNASELVVLIP-ELCRVTGLNAEMRSNFQLMRAMSSYTR 392
Db        427 MKLRYPHLP---------CLQVGQEHKHTYLPLEVCNIV---AGQRC-IKKLTDMQTSTM 473

Qy        393 MNPKQRT--DRLRAFNHRLQNTP-ESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGK 449
Db        474 IKATARSAPDREREINNLVKRADFNNDSYVQEFGLTISNSMMEVRGRVLPPPKLQY-GGR 532

Qy        450 VPAGENADWQRHFRDQR--MLTTPSDG--------------LDRWAV--IAPQRNSHE-- 489
Db        533 VSTGLTG--QQLFPPQNKVSLASPNQGVWDMRGKQFFTGVEIRIWAIACFAPQRTVREDA 590

Qy        490 LRTLLDSLYRAASGMGLRIRSPQEFIIY---DDRTGTYVRAMDDCVRSDPKLILCLVPND 546
Db        591 LRNFTQQLQKISNDAGMPIIGQPCFCKYATGPDQVEPMFRYL-KITFPGLQLVVVVLPG- 648

Qy        547 NAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLG---------YT 597
Db        649 KTPVYAEVKRVGDTVLGMATQCVQAKNV-NKTSPQTLSNLCLKINVKLGGINSILVPSIR 707

Qy        598 PWMIELPLSGLMTIGFDIA--KSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLA 655
Db        708 PKVFNEP---VIFLGADVTHPPAGDNKKPSIAAVVGSMDAHP-SRYAATVRVQQHRQEII 763

Qy        656 NTLWPMIAKALRQ-YQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIE---KLKTEY 711
Db        764 QELSSMVRELLIMFYKSTGGYKPHRIILYRDGVSEGQFPHVLQHELTAIREACIKLEPEY 823

Qy        712 ARVQLSPPQLAYIVVTRSMNTRFFL--------NGQNPPPGTIVDDVITLPERYDFYLVS 763
Db        824 R------PGITFIVVQKRHHTRLFCAEKKEQSGKSGNIPAGTTVDVGITHPTEFDFYLCS 877

Qy        764 QQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLA 821
Db        878 HQGIQGTSRPSHYHVLWDDNHFDSDELQCLTYQLCHTYVRCTRSVSIPAPAYYAHLVA 935
```

RESULT 11
ABB67045
ID ABB67045 standard; Protein; 984 AA.
XX
AC ABB67045;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 27927.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL11148.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 27927; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 984 AA;

Query Match 10.4%; Score 458.5; DB 22; Length 984;
Best Local Similarity 23.9%; Pred. No. 4.9e-33;
Matches 200; Conservative 144; Mismatches 343; Indels 151; Gaps 39;

```
Qy         91 GTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMGVLSN--HANLLGSGYL- 147
Db        142 GREGRPIVLRANHFQV-TMPRGYVHHYDINIQPDKCPRKVNREIIETMVHAYSKIFGVLK 200

Qy        148 --FDGLQLFTTR--------KFEQEITVLSGKSKLDIEYKISIKFVGFIS------CAEP 191
Db        201 PVFDGRNNLYTRDPLPIGNERLELEVT-LPGEGK-DRIFRVTIKWQAQVSLFNLEEALEG 258

Qy        192 RFLQV-------LNLILRRSMKGLNLELVGRNLF-DPRAKIEIREFKMELWPGYETSIRQ 243
Db        259 RTRQIPYDAILALDVVMRH-LPSMTYTPVGRSFFSSPEGYYHPLGGGREVWFGFHQSVRP 317

Qy        244 HEKDILLGTEITHKVM-RTETIYDIMRRCSHNPARHQDEVRVNVLDL------------- 289
Db        318 SQWKMMLNIDVSATAFYKAQPVIDFM--CEVLDIRDINEQRKPLTDSQRVKFTKEIKGLK 375

Qy        290 --IVLTDYNNRTYRINDV--------------DFGQTPKSTFSCKGRDISFVEYYLTKYN 333
```

```
Db        376 IEITHCGQMRRKYRVCNVTRRPAQMQSFPLQLENGQTVECTVA---------KYFLDKYR 426

Qy        334 IRIRDHNQPLLISKNRDKALKTNASELVVLIP-ELCRVTGLNAEMRSNFQLMRAMSSYTR 392

Db        427 MKLRYPHLP---------CLQVGQEHKHTYLPLEVCNIV---AGQRC-IKKLTDMQTSTM 473

Qy        393 MNPKQRT--DRLRAFNHRLQNTP-ESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGK 449

Db        474 IKATARSAPDREREINNLVKRADFNNDSYVQEFGLTISNSMMEVRGRVLPPPKLQY-GGR 532

Qy        450 VPAGENADWQRHFRDQR--MLTTPSDG--------------LDRWAV--IAPQRNSHE-- 489

Db        533 VSTGLTG--QQLFPPQNKVSLASPNQGVWDMRGKQFFTGVEIRIWAIACFAPQRTVREDA 590

Qy        490 LRTLLDSLYRAASGMGLRIRSPQEFIIY---DDRTGTYVRAMDDCVRSDPKLILCLVPND 546

Db        591 LRNFTQQLQKISNDAGMPIIGQPCFCKYATGPDQVEPMFRYL-KITFPGLQLVVVVLPG- 648

Qy        547 NAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLG---------YT 597

Db        649 KTPVYAEVKRVGDTVLGMATQCVQAKNV-NKTSPQTLSNLCLKINVKLGGINSILVPSIR 707

Qy        598 PWMIELPLSGLMTIGFDIA--KSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLA 655

Db        708 PKVFNEP---VIFLGADVTHPPAGDNKKPSIAAVVGSMDAHP-SRYAATVRVQQHRQEII 763

Qy        656 NTLWPMIAKALRQ-YQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIE---KLKTEY 711

Db        764 QELSSMVRELLIMFYKSTGGYKPHRIILYRDGVSEGQFPHVLQHELTAIREACIKLEPEY 823

Qy        712 ARVQLSPPQLAYIVVTRSMNTRFFL--------NGQNPPPGTIVDDVITLPERYDFYLVS 763

Db        824 R------PGITFIVVQKRHHTRLFCAEKKEQSGKSGNIPAGTTVDVGITHPTEFDFYLCS 877

Qy        764 QQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLA 821

Db        878 HQGIQGTSRPSHYHVLWDDNHFDSDELQCLTYQLCHTYVRCTRSVSIPAPAYYAHLVA 935


RESULT 12
AAM80292
ID   AAM80292 standard; Protein; 860 AA.
XX
AC   AAM80292;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human protein SEQ ID NO 3944.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US04098.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   01-SEP-2000; 2000US-0654936.
PR   15-SEP-2000; 2000US-0663561.
PR   20-OCT-2000; 2000US-0693325.
PR   30-NOV-2000; 2000US-0728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI   Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI   Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   N-PSDB; AAK53425.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities,
PT   useful in diagnosis and gene therapy -
XX
PS   Claim 20; Page 6203-6205; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC   (AAM80020) are omitted as the relevant pages from the sequence listing
CC   were missing at the time of publication.
XX
SQ   Sequence   860 AA;

  Query Match             10.2%;  Score 448;  DB 22;  Length 860;
  Best Local Similarity   22.9%;  Pred. No. 3.7e-32;
  Matches  201;  Conservative  145;  Mismatches  315;  Indels  218;  Gaps   39;

Qy         82 RPVELVSKK---GTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMGVLSN- 137

Db         12 QPLLMVPRRPGYGAMGKPIKLLANCFQVEI-PKIDVYLYEVDIKPDKCPRRVNREVVDSM 70

Qy        138 --HANLLGSG---YLFDGLQ-LFT-------TRKFEQEITVLSGKSKLDIEYKISIKFVG 184

Db         71 VQHFKVTIFGDRRPVYDGKRSLYTANPLPVATTGVDLDVT-LPGEGGKDRPFKVSIKFVS 129

Qy        185 FIS------------CAEPRFL---------QVLNLILRRSMKGLNLELVGRNLFD-PRA 222

Db        130 RVSWHLLHEVLTGRTLPEPLELDKPISTNPVHAVDVVLRH-LPSMKYTPVGRSFFSAPEG 188

Qy        223 KIEIREFKMELWPGYETSIRQHEKDILLGTEI---------------------------- 254

Db        189 YDHPLGGGREVWFGFHQSVRPAMWKMMLNIDVSATAFYKAQPVIQFMCEVLDIHNIDEQP 248

Qy        255 -----THKVMRTETIYDI---------MRR------CSHNPARHQDEVRVNVLDLIVLTD 294

Db        249 RPLTDSHRVKFTKEIKGLKVEVTHCGTMRRKYRVCNVTRRPASHQ--------------- 293

Qy        295 YNNRTYRINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALK 354

Db        294 ----TFPL-QLENGQTVERTVA---------QYFREKYTLQLKYPHLP---------CLQ 330

Qy        355 TNASELVVLIP-ELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTP 413

Db        331 VGQEQKHTYLPLEVCNI------------VAGQRCIKKLTDNQTSTMIKATARSAPDRQ 377

Qy        414 ESV-KVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKV-PA------GENADWQRHFRDQ 465

Db        378 EEISRLVRSANYETDPFVQEFQFKV--RDEMAHVTGRVLPAPMLQYGGRN---------- 425

Qy        466 RMLTTPSDG--------------LDRWAV--IAPQRNSHE--LRTLLDSLYRAASGMGLR 507

Db        426 RTVATPSHGVWDMRGKQFHTGVEIKMWAIACFATQRQCREEILKGFTDQLRKISKDAGMP 485

Qy        508 IRSPQEFIIY---DDRTGTYVRAMDDCVRSDPKLILCLVPNDNAERYSSIKKRGYVDRAV 564

Db        486 IQGQPCFCKYAQGADSVEPMFRHLKN-TYSGLQLIIVILPG-KTPVYAEVKRVGDTLLGM 543
```

```
Qy        565 PTQVVTLKTTKNRSLMSIATKIAIQLNCKLG---------YTPWMIELPLSGLMTIGFDI 615
Db        544 ATQCVQVKNVIKTSPQTL-SNLCLKINVKLGGINNILVPHQRPSVFQQP---VIFLGADV 599

Qy        616 A--KSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEH 673
Db        600 THPPAGDGKKPSIAAVVGSMDAHP-SRYCATVRVQRPRQEIIQDLASMVRELLIQFYKST 658

Qy        674 RKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIE---KLKTEYARVQLSPPQLAYIVVTRSM 730
Db        659 RFKPTRIIFYRDGVSEGQFRQVLYYELLAIREACISLEKDY------QPGITYIVVQKRH 712

Qy        731 NTRFFL--------NGQNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSS 782
Db        713 HTRLFCADRTERVGRSGNIPAGTTVDTDITHPYEFDFYLCSHAGIQGTSRPSHYHVLWDD 772

Qy        783 MGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLA 821
Db        773 NCFTADELQLLTYQLCHTYVRCTRSVSIPAPAYYAHLVA 811
```

RESULT 13
AAB94291
ID AAB94291 standard; Protein; 860 AA.
XX
AC AAB94291;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:14739.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 14739; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 860 AA;

Query Match 10.2%; Score 448; DB 22; Length 860;
Best Local Similarity 22.9%; Pred. No. 3.7e-32;
Matches 201; Conservative 145; Mismatches 315; Indels 218; Gaps 39;

```
Qy         82 RPVELVSKK---GTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMGVLSN- 137
Db         12 QPLLMVPRRPGYGAMGKPIKLLANCFQVEI-PKIDVYLYEVDIKPDKCPRRVNREVVDSM 70

Qy        138 --HANLLGSG---YLFDGLQ-LFT-------TRKFEQEITVLSGKSKLDIEYKISIKFVG 184
Db         71 VQHFKVTIFGDRRPVYDGKRSLYTANPLPVATTGVDLDVT-LPGEGGKDRPFKVSIKFVS 129

Qy        185 FIS-----------CAEPRFL---------QVLNLILRRSMKGLNLELVGRNLFD-PRA 222
Db        130 RVSWHLLHEVLTGRTLPEPLELDKPISTNPVHAVDVVLRH-LPSMKYTPVGRSFFSAPEG 188

Qy        223 KIEIREFKMELWPGYETSIRQHEKDILLGTEI---------------------------- 254
Db        189 YDHPLGGGREVWFGFHQSVRPAMWKMMLNIDVSATAFYKAQPVIQFMCEVLDIHNIDEQP 248

Qy        255 -----THKVMRTETIYDI---------MRR------CSHNPARHQDEVRVNVLDLIVLTD 294
Db        249 RPLTDSHRVKFTKEIKGLKVEVTHCGTMRRKYRVCNVTRRPASHQ--------------- 293

Qy        295 YNNRTYRINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALK 354
Db        294 ----TFPL-QLENGQTVERTVA---------QYFREKYTLQLKYPHLP---------CLQ 330

Qy        355 TNASELVVLIP-ELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTP 413
Db        331 VGQEQKHTYLPLEVCNI------------VAGQRCIKKLTDNQTSTMIKATARSAPDRQ 377

Qy        414 ESV-KVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKV-PA------GENADWQRHFRDQ 465
Db        378 EEISRLVRSANYETDPFVQEFQFKV--RDEMAHVTGRVLPAPMLQYGGRN---------- 425

Qy        466 RMLTTPSDG-------------LDRWAV--IAPQRNSHE--LRTLLDSLYRAASGMGLR 507
Db        426 RTVATPSHGVWDMRGKQFHTGVEIKMWAIACFATQRQCREEILKGFTDQLRKISKDAGMP 485

Qy        508 IRSPQEFIIY---DDRTGTYVRAMDDCVRSDPKLILCLVPNDNAERYSSIKKRGYVDRAV 564
Db        486 IQGQPCFCKYAQGADSVEPMFRHLKN-TYSGLQLIIVILPG-KTPVYAEVKRVGDTLLGM 543

Qy        565 PTQVVTLKTTKNRSLMSIATKIAIQLNCKLG---------YTPWMIELPLSGLMTIGFDI 615
Db        544 ATQCVQVKNVIKTSPQTL-SNLCLKINVKLGGINNILVPHQRPSVFQQP---VIFLGADV 599

Qy        616 A--KSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEH 673
Db        600 THPPAGDGKKPSIAAVVGSMDAHP-SRYCATVRVQRPRQEIIQDLASMVRELLIQFYKST 658

Qy        674 RKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIE---KLKTEYARVQLSPPQLAYIVVTRSM 730
Db        659 RFKPTRIIFYRDGVSEGQFRQVLYYELLAIREACISLEKDY------QPGITYIVVQKRH 712

Qy        731 NTRFFL--------NGQNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSS 782
Db        713 HTRLFCADRTERVGRSGNIPAGTTVDTDITHPYEFDFYLCSHAGIQGTSRPSHYHVLWDD 772
```

```
Qy        783 MGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLA 821
              : :::| |||::|| |   : :  :||   ||  :|
Db        773 NCFTADELQLLTYQLCHTYVRCTRSVSIPAPAYYAHLVA 811
```

RESULT 14
AAG51563
ID AAG51563 standard; Protein; 906 AA.
XX
AC AAG51563;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 65456.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.

```
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             9.9%;  Score 436.5;  DB 21;  Length 906;
  Best Local Similarity   23.3%;  Pred. No. 4.8e-31;
  Matches  197;  Conservative  139;  Mismatches  346;  Indels  163;  Gaps   37;

Qy         91 GTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMGVLSNHANLLGSGYL--- 147
Db         53 GTLGTKCIVKANHFLADL-PTKDLNQYDVTITPEVSSKSVNRAIIAELVRLYKESDLGRR 111

Qy        148 ---FDGLQL--------FTTRKFEQEIT-----VLSGKSKLDIEYKISIKFVG------- 184
Db        112 LPAYDGRKSLYTAGELPFTWKEFSVKIVDEDDGIING-PKRERSYKVAIKFVARANMHHL 170

Qy        185 --FIS-----CAEPRFLQVLNLILRR-SMKGLNLELVGRNLFDPRAKIEIREFK-MELWP 235
Db        171 GEFLAGKRADCPQ-EAVQILDIVLRELSVK--RFCPVGRSFFSPDIKTPQRLGEGLESWC 227

Qy        236 GYETSIRQHEKDI----------------------LLGTEITHKVMRTETIYDIMR--R 270
Db        228 GFYQSIRPTQMGLSLNIDMASAAFIEPLPVIEFVAQLLGKDVLSKPLSDSDRVKIKKGLR 287

Qy        271 CSHNPARHQDEVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTF----SCKGRDISFVE 326
Db        288 GVKVEVTHRANVR-------------RKYRVAGLTTQPTRELMFPVDENCTMK--SVIE 331

Qy        327 YYLTKYNIRIRDHNQPLLISKNRDKALKTNASELVVLIPELCR-VTGLNAEMRSNFQLMR 385
Db        332 YFQEMYGFTIQHTHLPCLQVGNQKKA--------SYLPMEACKIVEGQRYTKRLNEKQIT 383

Qy        386 AMSSYTRMNPKQR-TDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIV 444
Db        384 ALLKVTCQRPRDRENDILRTVQH---NAYDQDPYAKEFGMNISEKLASVEARILPAPWLK 440

Qy        445 FH-NGK----VPAGENADWQRHFRDQRMLTTPSDGLDRWAVIAPQRNSHE--LRTLLDSL 497
Db        441 YHENGKEKDCLP--QVGQW--NMMNKKMINGMT--VSRWACVNFSRSVQENVARGFCNEL 494

Qy        498 YRAASGMGLRIRSPQEFIIYDDRTG--------TYVRAMDDCVRSDPKLILCLVPNDNAE 549
Db        495 GQMCEVSGMEFNPEPVIPIYSARPDQVEKALKHVYHTSMNKTKGKELELLLAILPDNNGS 554

Qy        550 RYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLGYTPWMI------EL 603
Db        555 LYGDLKRICETELGLISQCCLTKHVFKISKQYLA-NVSLKINVKMGGRNTVLVDAISCRI 613

Qy        604 PL-SGLMTI--GFDIA--KSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTL 658
Db        614 PLVSDIPTIIFGADVTHPENGEESSPSIAAVVASQDWPEVTKYAGLVCAQAHRQELIQDL 673

Qy        659 WP-------------MIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIE 705
Db        674 YKTWQDPVRGTVSGGMIRDLLISFRKATGQKPLRIIFYRDGVSEGQFYQVLLYEL-DAIR 732

Qy        706 KLKTEYARVQLSP---PQLAYIVVTRSMNTRFFLNGQ----------NPPPGTIVDDVIT 752
Db        733 K-----ACASLEPNYQPPVTFIVVQKRHHTRLFANNHRDKNSTDRSGNILPGTVVDTKIC 787

Qy        753 LPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPA 812
Db        788 HPTEFDFYLCSHAGIQGTSRPAHYHVLWDENNFTADGIQSLTNNLCYTYARCTRSVSIVP 847

Qy        813 VCQYA 817
Db        848 PAYYA 852


RESULT 15
AAG51562
ID   AAG51562 standard; Protein; 983 AA.
XX
AC   AAG51562;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 65455.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
```

XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.

PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 9.9%; Score 436.5; DB 21; Length 983;
Best Local Similarity 23.3%; Pred. No. 5.5e-31;
Matches 197; Conservative 139; Mismatches 346; Indels 163; Gaps 37;

```
Qy        91 GTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMGVLSNHANLLGSGYL--- 147
             || |   ::: | |     |   :  | |   | : :  |   :::    |      |
Db       130 GTLGTKCIVKANHFLADL-PTKDLNQYDVTITPEVSSKSVNRAIIAELVRLYKESDLGRR 188

Qy       148 ---FDGLQL--------FTTRKFEQEIT-----VLSGKSKLDIEYKISIKFVG------- 184
               :|| :          || ::|  :|      :::|  | :  ||::||||
Db       189 LPAYDGRKSLYTAGELPFTWKEFSVKIVDEDDGIING-PKRERSYKVAIKFVARANMHHL 247

Qy       185 --FIS-----CAEPRFLQVLNLILRR-SMKGLNLELVGRNLFDPRAKIEIREFK-MELWP 235
               |:     | :   :|:|:::||  |:|      |||: | |  |  |  : :| |
Db       248 GEFLAGKRADCPQ-EAVQILDIVLRELSVK--RFCPVGRSFFSPDIKTPQRLGEGLESWC 304

Qy       236 GYETSIRQHEKDI----------------------LLGTEITHKVMRTETIYDIMR--R 270
             |:  |||  :  :                      ||| ::  | :       | :  |
Db       305 GFYQSIRPTQMGLSLNIDMASAAFIEPLPVIEFVAQLLGKDVLSKPLSDSDRVKIKKGLR 364

Qy       271 CSHNPARHQDEVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTF----SCKGRDISFVE 326
               |:  ||              | ||:  :    | :  |    :|  :  | :|
Db       365 GVKVEVETHRANVR-------------RKYRVAGLTTQPTRELMFPVDENCTMK--SVIE 408

Qy       327 YYLTKYNIRIRDHNQPLLISKNRDKALKTNASELVVLIPELCR-VTGLNAEMRSNFQLMR 385
             |:   |   |:  : | |   |: ||          |  | |: | |     | | : :
Db       409 YFQEMYGFTIQHTHLPCLQVGNQKKA--------SYLPMEACKIVEGQRYTKRLNEKQIT 460

Qy       386 AMSSYTRMNPKQR-TDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIV 444
             |:   |   |: |  | ||   |   |  :     ::: | : : :  |: ||:    :
Db       461 ALLKVTCQRPRDRENDILRTVQH---NAYDQDPYAKEFGMNISEKLASVEARILPAPWLK 517

Qy       445 FH-NGK----VPAGENADWQRHFRDQRMLTTPSDGLDRWAVIAPQRNSHE--LRTLLDSL 497
             :| |||    :|  :   |  :  :::|:   :  : |||  :   |:  |   |   : |
Db       518 YHENGKEKDCLP--QVGQW--NMMNKKMINGMT--VSRWACVNFSRSVQENVARGFCNEL 571

Qy       498 YRAASGMGLRIRSPQEFIIYDDRTG--------TYVRAMDDCVRSDPKLILCLVPNDNAE 549
               :    |:          ||  |           |  :|:     : :|:| ::|::|
Db       572 GQMCEVSGMEFNPEPVIPIYSARPDQVEKALKHVYHTSMNKTKGKELELLLAILPDNNGS 631

Qy       550 RYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLGYTPWMI------EL 603
               |  :|:    :  : :|    |     |   :|  :::::| |:|    ::      :
Db       632 LYGDLKRICETELGLISQCCLTKHVFKISKQYLA-NVSLKINVKMGGRNTVLVDAISCRI 690

Qy       604 PL-SGLMTI--GFDIA--KSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTL 658
             || | : ||  | |:   ::  :   :  |::|| |  : : |   |   :   |  |
Db       691 PLVSDIPTIIFGADVTHPENGEESSPSIAAVVASQDWPEVTKYAGLVCAQAHRQELIQDL 750

Qy       659 WP------------MIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIE 705
             :              ||   |  ::    : | ||:|||||||| |   |:  :|: | |
Db       751 YKTWQDPVRGTVSGGMIRDLLISFRKATGQKPLRIIFYRDGVSEGQFYQVLLYEL-DAIR 809

Qy       706 KLKTEYARVQLSP---PQLAYIVVTRSMNTRFFLNGQ----------NPPPGTIVDDVIT 752
             |     |  | |   | : :||| :  :|| | |            |  |||:||  |
Db       810 K-----ACASLEPNYQPPVTFIVVQKRHHTRLFANNHRDKNSTDRSGNILPGTVVDTKIC 864

Qy       753 LPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPA 812
             |  :|||| |    |||  |  |:||:     : : :| ||  :|: |   : :  :
Db       865 HPTEFDFYLCSHAGIQGTSRPAHYHVLWDENNFTADGIQSLTNNLCYTYARCTRSVSIVP 924

Qy       813 VCQYA 817
               ||
Db       925 PAYYA 929
```

Search completed: July 22, 2003, 07:01:02
Job time : 32.5674 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 06:59:33 ; Search time 11.1699 Seconds
(without alignments)
2220.565 Million cell updates/sec

Title: US-09-873-737A-2
Perfect score: 4389
Sequence: 1 MADDQGRGRRRPXNEDDSST.....VGTNLHSIPQNALEKKFYYL 843

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	108.5	2.5	1312	2	US-08-592-126-148	Sequence 148, App
2	108.5	2.5	1312	2	US-08-687-080-51	Sequence 51, Appl
3	107.5	2.4	634	4	US-09-134-001C-3390	Sequence 3390, Ap
4	107	2.4	912	4	US-09-134-001C-2993	Sequence 2993, Ap
5	106	2.4	3418	2	US-08-603-753D-4	Sequence 4, Appli
6	106	2.4	3418	3	US-08-755-587-44	Sequence 44, Appl
7	106	2.4	3418	4	US-09-099-753-4	Sequence 4, Appli
8	106	2.4	3418	4	US-08-986-106-4	Sequence 4, Appli
9	105.5	2.4	805	4	US-09-134-001C-4821	Sequence 4821, Ap
10	105.5	2.4	3457	2	US-08-416-603-4	Sequence 4, Appli
11	103	2.3	2329	3	US-08-755-587-16	Sequence 16, Appl
12	103	2.3	2548	4	US-09-172-422-1	Sequence 1, Appli
13	103	2.3	3418	2	US-08-639-501-2	Sequence 2, Appli
14	103	2.3	3418	3	US-09-044-946-2	Sequence 2, Appli
15	103	2.3	3418	3	US-09-044-908-2	Sequence 2, Appli
16	101.5	2.3	962	4	US-09-134-001C-4497	Sequence 4497, Ap
17	100	2.3	717	4	US-09-307-143-2	Sequence 2, Appli
18	100	2.3	1698	4	US-09-315-793-12	Sequence 12, Appl
19	99.5	2.3	645	2	US-08-592-126-144	Sequence 144, App
20	99.5	2.3	645	2	US-08-687-080-47	Sequence 47, Appl
21	99.5	2.3	1579	3	US-08-755-587-184	Sequence 184, App
22	99	2.3	425	3	US-09-109-204-2	Sequence 2, Appli
23	99	2.3	464	4	US-09-297-468-2	Sequence 2, Appli
24	99	2.3	1082	1	US-08-106-493A-2	Sequence 2, Appli
25	99	2.3	1082	1	US-08-429-264-2	Sequence 2, Appli
26	99	2.3	1139	1	US-08-832-883-2	Sequence 2, Appli
27	99	2.3	1139	2	US-08-832-877-2	Sequence 2, Appli
28	98.5	2.2	884	2	US-08-474-067-8	Sequence 8, Appli
29	98.5	2.2	884	2	US-08-474-068A-8	Sequence 8, Appli
30	98.5	2.2	884	2	US-08-472-481-7	Sequence 7, Appli
31	98.5	2.2	1183	2	US-08-447-031A-2	Sequence 2, Appli
32	98.5	2.2	1781	2	US-08-477-451-11	Sequence 11, Appl
33	98.5	2.2	2512	2	US-08-801-263A-9	Sequence 9, Appli
34	98.5	2.2	2512	3	US-09-102-248-9	Sequence 9, Appli
35	97.5	2.2	1041	4	US-09-134-001C-2949	Sequence 2949, Ap
36	96.5	2.2	830	1	US-07-977-434-6	Sequence 6, Appli
37	96.5	2.2	830	1	US-08-458-819-6	Sequence 6, Appli
38	96.5	2.2	830	5	PCT-US91-07035-6	Sequence 6, Appli
39	96.5	2.2	868	2	US-08-162-081B-33	Sequence 33, Appl
40	96.5	2.2	868	2	US-08-780-872-33	Sequence 33, Appl
41	96.5	2.2	868	4	US-09-085-957-33	Sequence 33, Appl
42	96	2.2	711	3	US-08-772-270A-12	Sequence 12, Appl
43	95.5	2.2	2802	4	US-09-542-331-1	Sequence 1, Appli
44	95.5	2.2	2802	4	US-09-510-791-1	Sequence 1, Appli
45	95	2.2	288	4	US-09-134-001C-5137	Sequence 5137, Ap

ALIGNMENTS

```
RESULT 1
US-08-592-126-148
; Sequence 148, Application US/08592126
; Patent No. 5821091
;  GENERAL INFORMATION:
;    APPLICANT:  Gregory Dolganov
;    TITLE OF INVENTION:  Transcripts Encoding Immunomodulatory
;    TITLE OF INVENTION:  Polypeptides
;    NUMBER OF SEQUENCES:  151
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/592,126
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sholtz, Charles K.
;      REGISTRATION NUMBER:  38,615
;      REFERENCE/DOCKET NUMBER:  4600-0111
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  148:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1312 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  Rad50.pro-translation of SEQ ID NO:54
US-08-592-126-148

  Query Match             2.5%;  Score 108.5;  DB 2;  Length 1312;
  Best Local Similarity   18.2%;  Pred. No. 0.32;
  Matches  105;  Conservative   93;  Mismatches  233;  Indels  145;  Gaps   22;

Qy        234 WPGYETSIRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARHQDEVRVNVLDLIVLT.293
              ||  |    : : | :     |  :   ||:  : :       :|      ::|  |
Db        164 WPLSEGKALKQKFDEIFSA--TRYIKALETLRQVRQTQGQKVKEYQ-------MELKYLK 214
```

```
Qy        294 DYNNRTYRINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKAL 353
Db        215 QYKEKACEIRD----------------QITSKEAQLTSSKEIVKSYENELDPLKNRLKEI 258

Qy        354 KTNASELVVLIPELCRVTGLNAEM-RSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNT 412
Db        259 EHNLSKIMKLDNEIKALDSRKKQMEKDNSELEEKMEKVFQGTDEQLND---LYHNHQRTV 315

Qy        413 PESVKVLRDWNMELDKNVTEVQGRIIGQQ--NIVFHNGKVPAGENAD-WQRHFRDQRML- 468
Db        316 REKERKLVDCHRELEK--LNKESRLLNQEKSELLVEQGRLQL--QADRHQEHIRARDSLI 371

Qy        469 -----TTPSDGLDRWAVIAPQ-RNSHEL----------------------RTL----LDS 496
Db        372 QSLATQLELDGFERGPFSERQIKNFHKLVRERQEGEAKTANQLMNDFAEKETLKQKQIDE 431

Qy        497 LYRAASGMGLRIRSPQEFIIYDDRTGTYVR--------AMDDCVRSDPKLILCLVPNDNA 548
Db        432 IRDKKTGLGRIIELKSEILSKKQNELKNVKYELQQLEGSSDRILELDQELIKAERELSKA 491

Qy        549 ERYSSIK-----------KRGYVDRAV----------------PTQVVTL---KTTKNRS 578
Db        492 EKNSNVETLKMEVISLQNEKADLDRTLRKLDQEMEQLNHHTTTRTQMEMLTKDKADKDEQ 551

Qy        579 LMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQN 638
Db        552 IRKIKSRHSDELTSLLGYFPNKKQLE-DWLHSKSKEI-NQTRDRLAKLNKELASSEQNKN 609

Qy        639 STYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEF 698
Db        610 ----------------------------HINNELKRREEQLSSYEDKLFDVCGSQDFES 640

Qy        699 EVKDIIEKLKTEYARVQLSPPQLAYIVVTRSMNTRF 734
Db        641 D----LDRLKEE---IEKSSKQRAMLAGATAVYSQF 669
```

RESULT 2
US-08-687-080-51
; Sequence 51, Application US/08687080
; Patent No. 5965427
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,080
; FILING DATE: 17-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/592,126
; FILING DATE: 26-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TRANS. OF RAD50 cDNA (SEQ. 54), NT.
; INDIVIDUAL ISOLATE: 389 TO 4324
US-08-687-080-51

Query Match 2.5%; Score 108.5; DB 2; Length 1312;
Best Local Similarity 18.2%; Pred. No. 0.32;
Matches 105; Conservative 93; Mismatches 233; Indels 145; Gaps 22;

```
Qy        234 WPGYETSIRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARHQDEVRVNVLDLIVLT 293
Db        164 WPLSEGKALKQKFDEIFSA--TRYIKALETLRQVRQTQGQKVKEYQ-------MELKYLK 214

Qy        294 DYNNRTYRINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKAL 353
Db        215 QYKEKACEIRD----------------QITSKEAQLTSSKEIVKSYENELDPLKNRLKEI 258

Qy        354 KTNASELVVLIPELCRVTGLNAEM-RSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNT 412
Db        259 EHNLSKIMKLDNEIKALDSRKKQMEKDNSELEEKMEKVFQGTDEQLND---LYHNHQRTV 315

Qy        413 PESVKVLRDWNMELDKNVTEVQGRIIGQQ--NIVFHNGKVPAGENAD-WQRHFRDQRML- 468
Db        316 REKERKLVDCHRELEK--LNKESRLLNQEKSELLVEQGRLQL--QADRHQEHIRARDSLI 371

Qy        469 -----TTPSDGLDRWAVIAPQ-RNSHEL----------------------RTL----LDS 496
Db        372 QSLATQLELDGFERGPFSERQIKNFHKLVRERQEGEAKTANQLMNDFAEKETLKQKQIDE 431

Qy        497 LYRAASGMGLRIRSPQEFIIYDDRTGTYVR--------AMDDCVRSDPKLILCLVPNDNA 548
Db        432 IRDKKTGLGRIIELKSEILSKKQNELKNVKYELQQLEGSSDRILELDQELIKAERELSKA 491

Qy        549 ERYSSIK-----------KRGYVDRAV----------------PTQVVTL---KTTKNRS 578
Db        492 EKNSNVETLKMEVISLQNEKADLDRTLRKLDQEMEQLNHHTTTRTQMEMLTKDKADKDEQ 551

Qy        579 LMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQN 638
Db        552 IRKIKSRHSDELTSLLGYFPNKKQLE-DWLHSKSKEI-NQTRDRLAKLNKELASSEQNKN 609

Qy        639 STYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEF 698
Db        610 ----------------------------HINNELKRREEQLSSYEDKLFDVCGSQDFES 640

Qy        699 EVKDIIEKLKTEYARVQLSPPQLAYIVVTRSMNTRF 734
Db        641 D----LDRLKEE---IEKSSKQRAMLAGATAVYSQF 669
```

RESULT 3
US-09-134-001C-3390
; Sequence 3390, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14

```
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3390
;   LENGTH: 634
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3390

  Query Match             2.4%;  Score 107.5;  DB 4;  Length 634;
  Best Local Similarity   20.4%;  Pred. No. 0.11;
  Matches  112;  Conservative   92;  Mismatches  177;  Indels  169;  Gaps   30;

Qy        160 EQEITVLSGKSKLDI-----------------EYKISIKFV-GFISCAE------PRFL 194
              |: ||:|:  : :|                   | :  :| | | :   :      |:
Db         73 ERGITILAKNTAIDYKGTRINILDTPGHADFGGEVERIMKMVDGVVLVVDAYEGTMPQTR 132

Qy        195 QVLNLILRRSMKGLNLELVGRNLFDPRAKIE-IREFKMELWPGYETSIRQHEKDIL---- 249
               ||   | :::| :   :|   :  | |: | : :  ::|:   | :  | :  ::
Db        133 FVLKKALEQNLKPV---VVVNKIDKPAARPEGVVDEVLDLFIELEANDEQLDFPVVYASA 189

Qy        250 ------LGTEITHKVMRT--ETIYDIMRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTYR 301
                   | :|   : |::  ||| | :      |  : ||    :   | | |||:   |
Db        190 VNGTASLDSEKQDENMQSLYETIIDYVPA----PVDNSDE---PLQFQIALLDYNDYVGR 242

Qy        302 INDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-------NRDKALK 354
              |     |:  :      | ::| ::   |  | |         ::|        |::  :
Db        243 IG---VGRVFRGKMRV-GDNVSLIKLDGTVKNFR---------VTKIFGYFGLKREEIEE 289

Qy        355 TNASELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTPE 414
               | :|:                |:|    :|     : :   :||     :
Db        290 AQAGDLI----------------------AVSGMEDINV---GETVTPHDHR-----D 317

Qy        415 SVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADW-----------QRHFR 463
               : |||     :|:   |:  ::          |    ||   |:            |:
Db        318 PLPVLR-----IDEPTLEMTFKV---------NNSPFAGREGDYVTARQIQERLDQQLET 363

Qy        464 DQRMLTTPSDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGT 523
              |  :  ||:|  | | |    |    |  |::::  |   |  |::  ||
Db        364 DVSLKVTPTDQPDSWVVAG--RGELHLSILIENMRR--EGFELQVSKPQ----------V 409

Qy        524 YVRAMDDCVRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIA 583
               :| :|  :    : : | ||::||   : |:  |    |   ::: : || |
Db        410 ILREIDGVLSEPFERVQCEVPSENAG--AVIESLG----ARKGEMLDMMTTDNG-----L 458

Qy        584 TKIAIQLNCK--LGYTPWMIELPLSGLMTIGFDIAKST----RDRKRAY------GALIA 631
              |::   :  :  :|||   :        || |: |   |    | | :|       ||||
Db        459 TRLIFMVPARGMIGYTTEFMS------MTRGYGIINHTFEEFRPRVKAQIGGRRNGALI- 511

Qy        632 SMDLQQNSTY 641
              |||  | ::|
Db        512 SMDQGQATSY 521


RESULT 4
US-09-134-001C-2993
; Sequence 2993, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2993
;   LENGTH: 912
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2993

  Query Match             2.4%;  Score 107;  DB 4;  Length 912;
  Best Local Similarity   18.1%;  Pred. No. 0.23;
  Matches  154;  Conservative  133;  Mismatches  313;  Indels  250;  Gaps   40;

Qy         85 ELVSKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMGVLSNHANLLGS 144
              :|:  | ||    :  : ::  ||||  ::   | :  |  |   :    | :  |:
Db        158 QLIGNKYTDNEYEVDNSRYYIEKTKP--FVIENKVYYEITFRNANDK---TSKYERLI-- 210

Qy        145 GYLFDGLQLFTTRKFE---QEITVLSGKSKLDI--EYKISIKFVGFISCAEPRFLQVLNL 199
              ||    :      ||    | |: |  |: : |  :|:|||:      |    | ::|:
Db        211 GYSKHRINTNYAVKFALEMQTISYLGIKANVIIINDYEISIR-----PCEFNNFAKILDY 265

Qy        200 ILR-----------RSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSIRQHEKD 247
               |:            : :|  || |:   | |        ||  ||    | |:
Db        266 DLKLQSNHNEYKKLMKLLKEFNLHLLDIVLLDD------DEFN-EL----ECSVNSESAV 314

Qy        248 ILLGTEITHKVMRTETIYDIMRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTYRIND--- 304
              | :      |:|    | :  :   |  |:   |  | :  : |    |:|::::
Db        315 INI-----FNVLRLARRYILTNKSGANVLRYLLFVFNNRIIRLQLPYNRNKCYKLSNLIL 369

Qy        305 ----VDFGQTPKSTFSCKGRD------ISFVEYYLTKYNIRIRDHNQPLLISKN--RDKA 352
                  | | |  | | || :      :  :||   :| : :|       | ||  ::|
Db        370 DYKCVPFDQMP-FTASLKGHNPNIYTLLQCIEYKGREYELLVRK------IQKNTLKNKK 422

Qy        353 LKTNASELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRT----DRLRAFNHR 408
              : |:  |    | :   |  |  :  :|         | :  | :      |:  : :
Db        423 VYTSKEE----IEQYGVVNELIDKYNNNL--------YYKHRPNREIHSFGDKYYLYENE 470

Qy        409 --LQNTPESVKVLRDWNME---------------LD-KNVTEVQGRIIGQQNIVFHNGK 449
               : :  :|:|:| : ::|               || |   |:  |:     |    |
Db        471 QSIMSIIKSIKLLSNESVEGYSNSVEFWLNNEYTSLDCKEKKEILLRMFSNSKISMVYGA 530

Qy        450 VPAGENA---DWQRHFRDQRMLTTPSDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGL 506
              |::        | |: ::          |||       | :|:: |      :
Db        531 AGTGKSTLINHICNFFYDKDVI----------VIA------NTNTAVDNIKRKIKLSNI 573

Qy        507 RIRSPQEFIIYD---------DRTGTYVRAMDDCVRSDPKLILCLVPNDNAE-------R 550
              :  :  :|:  |          |  ||       : :  : :  | |:  || :
Db        574 KTSTISKFLYNDKEKYDLLIIDEAGTVSNKDMNRILENKQFELLLIVGDNYQIESIDFGN 633

Qy        551 YSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMT 610
              :  | |   : : :  ::   :  |||  |:                  |
Db        634 WFEIAK-DVLSKNIINELTDMYRTKNDDLLYF----------------W---------- 665

Qy        611 IGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFD------------------ 652
                  || |::|      :|       |   :||  : | |:
Db        666 ------KSVREKKSNLNEII-------NMNKYSTRLDESIFNEFNKDEIILCLNYDGIYG 712

Qy        653 -------VLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSS--------GSLKQ--- 694
                    : ||      :   :::|     |:   |:|      |          ||: :
Db        713 INNINRLLQANNKNDSVIWGVKEY-----KVGDPILFNETNKYSPILFNNLKGSIIEIHV 767

Qy        695 -----LFEFEVKDIIEKLKTEYARVQL-SPPQLAYIVVTRSMNTRFFLNGQNPPPGTIVD 748
                  ||: |:  :| :|       : | |  : : ::  |   :    :| |       |
Db        768 FEEYILFDLEINKVINELDIISLEIDLISSSENSSVIRIRVEKS----DGLNDDDNDSSD 823

Qy        749 DVITLPERY----------DFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMC 798
               ::     |          :|  |   :        : |: |:::  : | ::
Db        824 SIVPFQVSYAISIHKAQGLEFNSVKIVISDDLDEQITNNIFYTAITRARENLK------- 876

Qy        799 HLYYNWSGTT 808
               :|  ||  |
Db        877 -IY--WSPRT 883


RESULT 5
US-08-603-753D-4
; Sequence 4, Application US/08603753D
; Patent No. 5891857
;  GENERAL INFORMATION:
;    APPLICANT:  HOLT, JEFFREY T.
;    APPLICANT:  JENSEN, ROY A.
```

```
;    APPLICANT:  PAGE, DAVID L.
;    APPLICANT:  KING, MARY-CLAIRE
;    APPLICANT:  SZABO, CSILLA I.
;    APPLICANT:  JETTON, THOMAS L.
;    APPLICANT:  ROBINSON-BENION, CHERYL L.
;    APPLICANT:  THOMPSON, MARILYN E.
;    TITLE OF INVENTION:  CHARACTERIZED BRCA1 AND BRCA2
;    TITLE OF INVENTION:  PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
;    TITLE OF INVENTION:  CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ARLES A. TAYLOR, JR.
;      STREET:  SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
;      STREET:  BOULEVARD
;      CITY:  DURHAM
;      STATE:  NORTH CAROLINA
;      COUNTRY:  USA
;      ZIP:  27707
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 800 kB storage
;      COMPUTER:  IBM PC/XT/AT compatible
;      OPERATING SYSTEM:  Windows 3.1
;      SOFTWARE:  WORD PERFECT 6.1 and ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/603,753D
;      FILING DATE:  20 FEB 1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.S. 08/373,799
;      FILING DATE:  17 JAN 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ARLES A. TAYLOR, JR.
;      REGISTRATION NUMBER:  39,395
;      REFERENCE/DOCKET NUMBER:  1242/2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 493-8000
;      TELEFAX:  (919) 419-0383
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3418
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens sapiens
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:  adult
;      TISSUE TYPE:  female breast
;      CELL TYPE:  normal breast tissue
;      CELL LINE:  HMEC
;      ORGANELLE:  no
;    FEATURE:
;      NAME/KEY:  BRCA2 protein
;      LOCATION:  1 to 3418; Genbank locus HSU43746
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:  BRCA2 protein has a negative
;      OTHER INFORMATION:  regulatory effect on growth of human mammary cells.
;    PUBLICATION INFORMATION:
;      AUTHORS:  Wooster, R. et al.
;      TITLE:  Identification of the breast cancer
;      TITLE:  susceptability gene BRCA2
;      JOURNAL:  Nature
;      VOLUME:  379
;      PAGES:  789-792
;      DATE:  1995
;      RELEVANT RESIDUES IN SEQ ID NO:  4:  granin box
;      RELEVANT RESIDUES IN SEQ ID NO:  domain at amino acids 3334-3344
US-08-603-753D-4
```

```
  Query Match             2.4%;  Score 106;  DB 2;  Length 3418;
  Best Local Similarity   22.1%;  Pred. No. 3.3;
  Matches   93;  Conservative   46;  Mismatches  162;  Indels  120;  Gaps   19;

Qy         73 GDQYDYLNTRPVELVSKKGTDGVP----VMLQTNFFRXKTKPEWRIVHYHV--------- 119
              | |:::   |    : :| |  ||     :|:|      |  | |   ||
Db       1122 GSQFEFTQFRKPSYILQKSTFEVPENQMTILKT------TSEECRDADLHVIMNAPSIGQ 1175

Qy        120 -----EFVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKS---- 170
                  :|  |:|  |   |:| |  |   ||| |          | |:      |
Db       1176 VDSSKQFEGTVEIKRKFAGLLKNDCNKSASGYLTD----------ENEVGFRGFYSAHGT 1225

Qy        171 KLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFK 230
              ||::  :   | |   |  |       |:    | :   : |      |      :  ||
Db       1226 KLNVSTEALQKAVKLFSDIE-------NISEETSAEVHPISLSSSKCHDS----VVSMFK 1274

Qy        231 MELWPGYETS--------IRQHEKDILLGT---EITHKVMRT----ETIYDIMRRCSHN- 274
              :|       |        | |:  ::  ||   |||    |     :  |    | |||
Db       1275 IENHNDKTVSEKNNKCQLILQNNIEMTTGTFVEEITENYKRNTENEDNKYTAASRNSHNL 1334

Qy        275 ------------PARHQDEVRVNVLDLIVLTDYNNRTYR---------------INDVD 306
                         |:||     ||:  || :|   :                   ::|:
Db       1335 EFDGSDSSKNDTVCIHKDET-----DLL-FTDQHNICLKLSGQFMKEGNTQIKEDLSDLT 1388

Qy        307 FGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKT----NASELVV 362
              | :  |:  :| |   :  :   ||    |:|       :  |   :|    | |
Db       1389 FLEVAKAQEACHGNTSNKEQLTATKTEQNIKDF-------ETSDTFFQTASGKNISVAKE 1441

Qy        363 LIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPK-----QRTDRLRAFNHRLQNTPESVK 417
              |  ::        |   || |   : |  | |       : || ::   |::    |||
Db       1442 LFNKIVNFFDQKPEELHNFSLNSELHSDIRKNKMDILSYEETDIVK---HKI--LKESVP 1496

Qy        418 V 418
              |
Db       1497 V 1497
```

```
RESULT 6
US-08-755-587-44
; Sequence 44, Application US/08755587
; Patent No. 6045997
;  GENERAL INFORMATION:
;    APPLICANT:  Futreal, Phillip A
;    APPLICANT:  Wooster, Richard F
;    APPLICANT:  Ashworth, Alan
;    APPLICANT:  Stratton, Michael R
;    TITLE OF INVENTION:  Materials and methods relating to the
;    TITLE OF INVENTION:  identification and sequencing of the BRCA2 cancer
;    TITLE OF INVENTION:  susceptibility gene and uses thereof.
;    NUMBER OF SEQUENCES:  222
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson
;      STREET:  310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
;      CITY:  Raleigh
;      STATE:  NC
;      COUNTRY:  USA
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/755,587
;      FILING DATE:  25-NOV-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9523959.6
;      FILING DATE:  23-NOV-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9525555.0
;      FILING DATE:  14-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9617961.9
;      FILING DATE:  28-AUG-1996
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kenneth D Sibley
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5405-135
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3418 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-755-587-44

  Query Match             2.4%;  Score 106;  DB 3;  Length 3418;
  Best Local Similarity   22.1%;  Pred. No. 3.3;
  Matches   93;  Conservative   46;  Mismatches  162;  Indels  120;  Gaps   19;

Qy         73 GDQYDYLNTRPVELVSKKGTDGVP----VMLQTNFFRXKTKPEWRIVHYHV--------- 119
              | |:::   |    : :| |  ||     :|:|      |  | |    ||
Db       1122 GSQFEFTQFRKPSYILQKSTFEVPENQMTILKT------TSEECRDADLHVIMNAPSIGQ 1175

Qy         120 -----EFVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKS---- 170
                  :|  |:|  |   |:| |  |   |||| |          | |:      |
Db       1176 VDSSKQFEGTVEIKRKFAGLLKNDCNKSASGYLTD----------ENEVGFRGFYSAHGT 1225

Qy         171 KLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFK 230
               ||::  :   | |   |  |       |:    | :   : |      |      :  ||
Db       1226 KLNVSTEALQKAVKLFSDIE-------NISEETSAEVHPISLSSSKCHDS----VVSMFK 1274

Qy         231 MELWPGYETS--------IRQHEKDILLGT---EITHKVMRT----ETIYDIMRRCSHN- 274
               :|      |        | |:  ::  ||   |||    |     : |    | |||
Db       1275 IENHNDKTVSEKNNKCQLILQNNIEMTTGTFVEEITENYKRNTENEDNKYTAASRNSHNL 1334

Qy         275 ------------PARHQDEVRVNVLDLIVLTDYNNRTYR----------------INDVD 306
                         |:||      ||:  || :|   :                    ::|:
Db       1335 EFDGSDSSKNDTVCIHKDET-----DLL-FTDQHNICLKLSGQFMKEGNTQIKEDLSDLT 1388

Qy         307 FGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKT----NASELVV 362
               | :  |:  :| |   :  :   ||    |:|      :  |   :|    | |
Db       1389 FLEVAKAQEACHGNTSNKEQLTATKTEQNIKDF-------ETSDTFFQTASGKNISVAKE 1441

Qy         363 LIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPK-----QRTDRLRAFNHRLQNTPESVK 417
               |  ::        |   || |   : |  | |       : || ::   |::    |||
Db       1442 LFNKIVNFFDQKPEELHNFSLNSELHSDIRKNKMDILSYEETDIVK---HKI--LKESVP 1496

Qy         418 V 418
               |
Db       1497 V 1497


RESULT 7
US-09-099-753-4
; Sequence 4, Application US/09099753
; Patent No. 6149903
;  GENERAL INFORMATION:
;    APPLICANT:  HOLT, JEFFREY T.
;    APPLICANT:  JENSEN, ROY A.
;    APPLICANT:  PAGE, DAVID L.
;    APPLICANT:  KING, MARY-CLAIRE
;    APPLICANT:  SZABO, CSILLA I.
;    APPLICANT:  JETTON, THOMAS L.
;    APPLICANT:  ROBINSON-BENION, CHERYL L.
;    APPLICANT:  THOMPSON, MARILYN E.
;    TITLE OF INVENTION:  CHARACTERIZED BRCA1 AND BRCA2
;    TITLE OF INVENTION:  PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
;    TITLE OF INVENTION:  CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ARLES A. TAYLOR, JR.
;      STREET:  SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
;      STREET:  BOULEVARD
;      CITY:  DURHAM
;      STATE:  NORTH CAROLINA
;      COUNTRY:  USA
;      ZIP:  27707
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 800 kB storage
;      COMPUTER:  IBM PC/XT/AT compatible
;      OPERATING SYSTEM:  Windows 3.1
;      SOFTWARE:  WORD PERFECT 6.1 and ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/099,753
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/603,753
;      FILING DATE:  20 FEB 1996
;      APPLICATION NUMBER:  U.S. 08/373,799
;      FILING DATE:  17 JAN 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ARLES A. TAYLOR, JR.
;      REGISTRATION NUMBER:  39,395
;      REFERENCE/DOCKET NUMBER:  1242/2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 493-8000
;      TELEFAX:  (919) 419-0383
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3418
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens sapiens
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:  adult
;      TISSUE TYPE:  female breast
;      CELL TYPE:  normal breast tissue
;      CELL LINE:  HMEC
;      ORGANELLE:  no
;    FEATURE:
;      NAME/KEY:  BRCA2 protein
;      LOCATION:  1 to 3418; Genbank locus HSU43746
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:  BRCA2 protein has a negative
;      OTHER INFORMATION:  regulatory effect on growth of human mammary cells.
;    PUBLICATION INFORMATION:
;      AUTHORS:  Wooster, R. et al.
;      TITLE:  Identification of the breast cancer
;      TITLE:  susceptability gene BRCA2
;      JOURNAL:  Nature
;      VOLUME:  379
;      PAGES:  789-792
;      DATE:  1995
;      RELEVANT RESIDUES IN SEQ ID NO:  4:  granin box
;      RELEVANT RESIDUES IN SEQ ID NO:  domain at amino acids 3334-3344
US-09-099-753-4

  Query Match             2.4%;  Score 106;  DB 4;  Length 3418;
  Best Local Similarity   22.1%;  Pred. No. 3.3;
  Matches   93;  Conservative   46;  Mismatches  162;  Indels  120;  Gaps   19;

Qy         73 GDQYDYLNTRPVELVSKKGTDGVP----VMLQTNFFRXKTKPEWRIVHYHV--------- 119
              | |:::   |    : :| |  ||     :|:|      |  | |    ||
Db       1122 GSQFEFTQFRKPSYILQKSTFEVPENQMTILKT------TSEECRDADLHVIMNAPSIGQ 1175

Qy         120 -----EFVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKS---- 170
                  :|  |:|  |   |:| |  |   |||| |          | |:      |
Db       1176 VDSSKQFEGTVEIKRKFAGLLKNDCNKSASGYLTD----------ENEVGFRGFYSAHGT 1225

Qy         171 KLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFK 230
               ||::  :   | |   |  |       |:    | :   : |      |      :  ||
Db       1226 KLNVSTEALQKAVKLFSDIE-------NISEETSAEVHPISLSSSKCHDS----VVSMFK 1274
```

```
Qy        231 MELWPGYETS--------IRQHEKDILLGT---EITHKVMRT----ETIYDIMRRCSHN- 274
              :|        |        | |:  ::  ||   |||    |     :  |    | |||
Db       1275 IENHNDKTVSEKNNKCQLILQNNIEMTTGTFVEEITENYKRNTENEDNKYTAASRNSHNL 1334

Qy        275 ------------PARHQDEVRVNVLDLIVLTDYNNRTYR---------------INDVD 306
                          |:||      ||:  || :|   :                   ::|:
Db       1335 EFDGSDSSKNDTVCIHKDET-----DLL-FTDQHNICLKLSGQFMKEGNTQIKEDLSDLT 1388

Qy        307 FGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKT----NASELVV 362
              | :  |:  :| |   :  :   ||    |:|        :  |   :|    | |
Db       1389 FLEVAKAQEACHGNTSNKEQLTATKTEQNIKDF-------ETSDTFFQTASGKNISVAKE 1441

Qy        363 LIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPK-----QRTDRLRAFNHRLQNTPESVK 417
              |  ::        |   || |   : |  | |       : || ::   |::    |||
Db       1442 LFNKIVNFFDQKPEELHNFSLNSELHSDIRKNKMDILSYEETDIVK---HKI--LKESVP 1496

Qy        418 V 418
              |
Db       1497 V 1497
```

```
RESULT 8
US-08-986-106-4
; Sequence 4, Application US/08986106
; Patent No. 6177410
;  GENERAL INFORMATION:
;    APPLICANT:  HOLT, JEFFREY T.
;    APPLICANT:  JENSEN, ROY A.
;    APPLICANT:  KING, MARY-CLAIRE
;    APPLICANT:  STEINER, MITCHELL S.
;    APPLICANT:  ROBINSON-BENION, CHERYL L.
;    APPLICANT:  THOMPSON, MARILYN E.
;    TITLE OF INVENTION:  THERAPEUTIC METHODS FOR
;    TITLE OF INVENTION:  PROSTATE CANCER
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ARLES A. TAYLOR, JR.
;      STREET:  SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
;      STREET:  BOULEVARD
;      CITY:  DURHAM
;      STATE:  NORTH CAROLINA
;      COUNTRY:  USA
;      ZIP:  27707
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44MB storage
;      COMPUTER:  IBM PC/XT/AT compatible
;      OPERATING SYSTEM:  Windows 3.1
;      SOFTWARE:  WORD PERFECT 6.1 and ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/986,106
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/603,753
;      FILING DATE:  20 FEB 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ARLES A. TAYLOR, JR.
;      REGISTRATION NUMBER:  39,395
;      REFERENCE/DOCKET NUMBER:  1242/3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 493-8000
;      TELEFAX:  (919) 419-0383
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3418
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    FEATURE:
;      NAME/KEY:  BRCA2 protein
;    PUBLICATION INFORMATION:
;      AUTHORS:  Wooster, R. et al.
;      TITLE:  Identification of the breast cancer
;      TITLE:  susceptability gene BRCA2
;      JOURNAL:  Nature
;      VOLUME:  379
;      PAGES:  789-792
;      DATE:  1995
;      RELEVANT RESIDUES IN SEQ ID NO:  4:  granin box domain at
;      RELEVANT RESIDUES IN SEQ ID NO:  amino acids 3334-3344
US-08-986-106-4

  Query Match             2.4%;  Score 106;  DB 4;  Length 3418;
  Best Local Similarity   22.1%;  Pred. No. 3.3;
  Matches   93;  Conservative   46;  Mismatches  162;  Indels  120;  Gaps   19;

Qy         73 GDQYDYLNTRPVELVSKKGTDGVP----VMLQTNFFRXKTKPEWRIVHYHV--------- 119
              | |:::   |    : :| |  ||     :|:|      |  | |   ||
Db       1122 GSQFEFTQFRKPSYILQKSTFEVPENQMTILKT------TSEECRDADLHVIMNAPSIGQ 1175

Qy        120 -----EFVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKS---- 170
                  :|  |:|  |   |:| |  |   |||| |          | |:      |
Db       1176 VDSSKQFEGTVEIKRKFAGLLKNDCNKSASGYLTD----------ENEVGFRGFYSAHGT 1225

Qy        171 KLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFK 230
              ||::  :   | |   |  |       |:    | :   : |     |      : ||
Db       1226 KLNVSTEALQKAVKLFSDIE-------NISEETSAEVHPISLSSSKCHDS----VVSMFK 1274

Qy        231 MELWPGYETS--------IRQHEKDILLGT---EITHKVMRT----ETIYDIMRRCSHN- 274
              :|        |        | |:  ::  ||   |||    |     :  |    | |||
Db       1275 IENHNDKTVSEKNNKCQLILQNNIEMTTGTFVEEITENYKRNTENEDNKYTAASRNSHNL 1334

Qy        275 ------------PARHQDEVRVNVLDLIVLTDYNNRTYR---------------INDVD 306
                          |:||      ||:  || :|   :                   ::|:
Db       1335 EFDGSDSSKNDTVCIHKDET-----DLL-FTDQHNICLKLSGQFMKEGNTQIKEDLSDLT 1388

Qy        307 FGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKT----NASELVV 362
              | :  |:  :| |   :  :   ||    |:|        :  |   :|    | |
Db       1389 FLEVAKAQEACHGNTSNKEQLTATKTEQNIKDF-------ETSDTFFQTASGKNISVAKE 1441

Qy        363 LIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPK-----QRTDRLRAFNHRLQNTPESVK 417
              |  ::        |   || |   : |  | |       : || ::   |::    |||
Db       1442 LFNKIVNFFDQKPEELHNFSLNSELHSDIRKNKMDILSYEETDIVK---HKI--LKESVP 1496

Qy        418 V 418
              |
Db       1497 V 1497 .
```

```
RESULT 9
US-09-134-001C-4821
; Sequence 4821, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4821
;   LENGTH: 805
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4821

  Query Match             2.4%;  Score 105.5;  DB 4;  Length 805;
  Best Local Similarity   22.6%;  Pred. No. 0.26;
  Matches   51;  Conservative   36;  Mismatches   62;  Indels   77;  Gaps   11;
```

```
Qy        220 PRAKIEIREFKMELWPGYETSIRQ--HEK----------DILLGTEITHK---------- 257
              | |:|     :|::    :|: |:  |||          ||||||::  |
Db        577 PHARI----IRMDV----DTTSRKGAHEKLLNDFEAGKGDILLGTQMIAKGLDYPNITLV 628

Qy        258 -VMRTETI-----------YDIMRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTYRIND 304
               |:  :|:           | :: : |    ||: | :|      ::  ||   | | |
Db        629 GVLNADTMLNLPDFRASERTYQLLTQVSGRAGRHEKEGQV------IIQTYNPDHYSIKD 682

Qy        305 VDFGQTPKSTFSCKGRDISFVE-------------YYLTKYNIRIRDHNQPLLISKNRD 350
              |               :|| :              |:|  : |   |  : :: ||:
Db        683 VKLNDY----------LSFYQKEMNYRKLGKYPPYYFLINFTISHTDIKKVMMASKHIH 731

Qy        351 KALKTNASE-LVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNP 395
              : |  : ||   || |    :  :| | |  ||::    |  :::|
Db        732 QILVQHLSEKAFVLGPSPAALARINNEYR--FQILVKYKSEPQLHP 775


RESULT 10
US-08-416-603-4
; Sequence 4, Application US/08416603
; Patent No. 5866780
;  GENERAL INFORMATION:
;    APPLICANT:  Law, Marcus
;    APPLICANT:  Hebara, Ledare
;    APPLICANT:  Reddick, Bradford B.
;    TITLE OF INVENTION:  Maize Chlorotic Dwarf Virus Genome and
;    TITLE OF INVENTION:  Uses Therefor
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/416,603
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lloyd, Jeffrey
;      REGISTRATION NUMBER:  35,589
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  904-375-8100
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3457 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-416-603-4

  Query Match             2.4%;  Score 105.5;  DB 2;  Length 3457;
  Best Local Similarity   20.0%;  Pred. No. 3.8;
  Matches  141;  Conservative   96;  Mismatches  221;  Indels  247;  Gaps   39;

Qy        169 KSKLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKGLNLE------------LVGR- 215
              |: |: |  |:: |   ::|   || |:|:    |  | ||:             |:||
Db       2264 KAALETEEYIALLFA--VACEAGRFTQILDKPPSR-RKILNMSERYNAYIEQEKGLIGRL 2320

Qy        216 --------------NLFDPRAKIEIREFKM-ELWPGYETSIRQHEKDILLGTEI--THKV 258
                           :|   | | :  ||:   :   |  : : | || |  |:  :|:
Db       2321 SKPAKICLAIGTGVAIFGALAGIGVGLFKLIAHFNKDEEEVDEIEFDI-LSPEMSGSHES 2379

Qy        259 MRTETIYDIMRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTYRINDVDFGQTP------- 311
               :  | |    |    ||| | |                    |: |   |
Db       2380 GQHTTRYVTKERVPSKPARRQHEF--------------------DLMFDNLPTPQVEEL 2418

Qy        312 KSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALK----TNASE------LV 361
              ||  :|     |  : : |:|   ::    |  :|| :| ::      :||:      |
Db       2419 KSEMTC----ASASDEHKTQY---VKRRVGP--VSKRKDASVAEISGAHASDQHHTEYLK 2469

Qy        362 VLIPELCR-------VTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHR--LQNT 412
              :| : |        |   : |  |:  |:| :          :||   ||      |::
Db       2470 ARVPLMKRIATKESYVVTYDDEPSSHISLVRRI---------RRTRLARAIKQMAVLEDF 2520

Qy        413 PESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPS 472
              | ::: :| |     :|    :| |: : :   :|| :|   | :   ::  ||
Db       2521 PSTLEEIRLWR----QNAAN-KGVIVPKYST---SGKFFSGLLDDEEEEPQNVNMLNEED 2572

Qy        473 DGLDR---------WAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGT 523
              :|:           :|| |::|  ||  :::         |:  :  ::
Db       2573 IEVDKRMFEKISEVISVIQPRKN--ELERMIEE--------GVHHKVVKQ---------- 2612

Qy        524 YVRAMDDCVRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIA 583
              |  |  :  || ::  |   |     |::             :| |  |: |: |::
Db       2613 -ARVNDKGLAKDPNMVTIL--TDKLINISAV------------IVNLTPTR-RAYMNVV 2655

Qy        584 TKIAIQLNCKLGYTPWMIELPLSG--LMTIGFD-IAKSTRDRKRAYGALIASMDLQQNST 640
              |   : |   |    :|    |  |  | |  : | | |  |                |
Db       2656 RLIGTIVVCPAHY----LEALEEGDELYFICFSLVIKLTFDPSRV-------------T 2697

Qy        641 YFSTVTECSAFDVLANTLWPMI-----AKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQL 695
              ::  :   :| | | : | |        |  : |           ::||  :
Db       2698 LVNSQQDLMVWD-LGNMVPPSIDTLKMIPTLEDWDH----------FQDGPGA------ 2739

Qy        696 FEFEVKDIIEKLKTEYARVQLSPPQLAYIVVTRSMNTRFFLNGQNPPPGTIVDDVITLPE 755
              | |     |  | |                   :||                   :|: |
Db       2740 --FAVTKYNSKFPTNY-----------------INT-----------------LTMIE 2761

Qy        756 RYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHL 800
              |         :|  | :||     || ||        | |:|  |
Db       2762 R---------IRANTQNPTG---CYSMMGSQHTITTGLRYQMFSL 2794


RESULT 11
US-08-755-587-16
; Sequence 16, Application US/08755587
; Patent No. 6045997
;  GENERAL INFORMATION:
;    APPLICANT:  Futreal, Phillip A
;    APPLICANT:  Wooster, Richard F
;    APPLICANT:  Ashworth, Alan
;    APPLICANT:  Stratton, Michael R
;    TITLE OF INVENTION:  Materials and methods relating to the
;    TITLE OF INVENTION:  identification and sequencing of the BRCA2 cancer
;    TITLE OF INVENTION:  susceptibility gene and uses thereof.
;    NUMBER OF SEQUENCES:  222
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson
;      STREET:  310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
;      CITY:  Raleigh
;      STATE:  NC
;      COUNTRY:  USA
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/755,587
;      FILING DATE:  25-NOV-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9523959.6
;      FILING DATE:  23-NOV-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9525555.0
;      FILING DATE:  14-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9617961.9
;      FILING DATE:  28-AUG-1996
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kenneth D Sibley
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5405-135
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2329 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-755-587-16

  Query Match             2.3%;  Score 103;  DB 3;  Length 2329;
  Best Local Similarity   22.1%;  Pred. No. 3.3;
  Matches   85;  Conservative   44;  Mismatches  147;  Indels  108;  Gaps   17;

Qy         73 GDQYDYLNTRPVELVSKKGTDGVP----VMLQTNFFRXKTKPEWRIVHYHV--------- 119
              | |:::   |    : :| |  ||     :|:|      |  | |    ||
Db        841 GSQFEFTQFRKPSYILQKSTFEVPENQMTILKT------TSEECRDADLHVIMNAPSIGQ 894

Qy        120 -----EFVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKS---- 170
                  :|  |:|  |   |:| |  |   ||| |           | |:      |
Db        895 VDSSKQFEGTVEIKRKFAGLLKNDCNKSASGYLTD----------ENEVGFRGFYSAHGT 944

Qy        171 KLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFK 230
              ||::  :  | |  |  |  .  |:    | :  : |      |      :  ||
Db        945 KLNVSTEALQKAVKLFSDIE-------NISEETSAEVHPISLSSSKCHDS----VVSMFK 993

Qy        231 MELWPGYETS--------IRQHEKDILLGT---EITHKVMRT----ETIYDIMRRCSHN- 274
              :|       |        | |:  ::  ||   |||    |     :  |    | |||
Db        994 IENHNDKTVSEKNNKCQLILQNNIEMTTGTFVEEITENYKRNTENEDNKYTAASRNSHNL 1053

Qy        275 ------------PARHQDEVRVNVLDLIVLTDYNNRTYR----------------INDVD 306
                         |:||      ||:  || :|   :                  ::|:
Db       1054 EFDGSDSSKNDTVCIHKDET-----DLL-FTDQHNICLKLSGQFMKEGNTQIKEDLSDLT 1107

Qy        307 FGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQ-----PLLISKNRDKALKTNASELV 361
              | :  |:  :| |   :  :   ||    |:|              ||   | | : :::|
Db       1108 FLEVAKAQEACHGNTSNKEQLTATKTEQNIKDFETSDTFFQTASGKNISVA-KESFNKIV 1166

Qy        362 VLI---PELCRVTGLNAEMRSNFQ 382
                 ||      ||:|: |: :
Db       1167 NFFDQKPEELHNFSLNSELHSDIR 1190


RESULT 12
US-09-172-422-1
; Sequence 1, Application US/09172422A
; Patent No. 6300485
; GENERAL INFORMATION:
;  APPLICANT: Adams, Arwen E.
;  APPLICANT: Chiu, Choi Ying
;  APPLICANT: Duhl, David
;  APPLICANT: Gorman, Susan W.
;  APPLICANT: Leng, Song
;  APPLICANT: Sheffield, Val
;  APPLICANT: Welch, Juliet
;  TITLE OF INVENTION: MYOSIN IXA AND CYCLIC NUCLEOTIDE GATED
;  TITLE OF INVENTION: CHANNEL-15 (CNGC-15) POLYNUCLEOTIDES, POLYPEPTIDES,
;  TITLE OF INVENTION: COMPOSITIONS, METHODS, AND USES THEREOF
;  FILE REFERENCE: 200130.442
;  CURRENT APPLICATION NUMBER: US/09/172,422A
;  CURRENT FILING DATE: 1998-10-14
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 2548
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-09-172-422-1

  Query Match             2.3%;  Score 103;  DB 4;  Length 2548;
  Best Local Similarity   17.8%;  Pred. No. 3.9;

  Matches  161;  Conservative  121;  Mismatches  319;  Indels  304;  Gaps   38;

Qy         77 DYL-NTRPVELVSKKGTDGVPVMLQTNFFRXKTKPEW--RIVHYH-----VEFVPTIENP 128
              ||: ||  : |:||| |  : :: : : |   |       :  | |     :|| | :  |
Db        588 DYIDNTCCINLISKKPTGLLHLLDEESNFPQATNQTLLDKFKHQHEDNSYIEF-PAVMEP 646

Qy        129 ---------RVRMGV------------------LSNHANLLGSGYL-FDGLQLF---TTR 157
                      :|: ||                  | :  |   || :  | : :|     |
Db        647 AFIIKHYAGKVKYGVKDFREKNTDHMRPDIVALLRSSKNAFISGMIGIDPVAVFRWAILR 706

Qy        158 KFEQEITVLSGKSKLDIEYKI-----------SIKFVGFIS-CAEPRFLQVLN------ 198
              | : :       | :|  |             |:    |:       | |::|
Db        707 AFFRAMVAFREAGKRNIHRKTGHDDTAPCAILKSMDSFSFLQHPVHQRSLEILQRCKEEK 766

Qy        199 -LILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSIRQH---------EKDI 248
               | |:: :    :| | |  : : : :  :     | |  | |||          :||
Db        767 YSITRKNPRTPLSDLQGMNALNEKNQHDTFDI---AWNG-RTGIRQSRLSSGTSLLDKDG 822

Qy        249 LLGTEITHKVMRTETIYDIMRRCSH-------------NPARH------QDEVRVNVLD 288
              :     : |::  |  : |: |  :               |  :|      || :  ::|
Db        823 IFANSTSSKLL--ERAHGILTRNKNFKSKPALPKHLLEVNSLKHLTRLTLQDRITKSLLH 880

Qy        289 L--------------------------------IVLTDYNNRTYRINDV---------- 305
              |                                  : ::      | :||
Db        881 LHKKKKPPSISAQFQASLSKLMETLGQAEPYFVKCIRSNAEKLPLRFSDVLVLRQLRYTG 940

Qy        306 --DFGQTPKSTFSCKGRDISFVEYY--LTKYNI-----RIRDHNQPLLISKNRDKALKTN 356
              :  |  :| :| |     || ::  |   ||       |:|  : : :: :  :  ||
Db        941 MLETVQIRQSGYSSKYSFQDFVSHFHVLLPRNIIPSKFNIQDFFRKINLNPDNYQVGKT- 999

Qy        357 ASELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDR-LRAFNHRLQNTPES 415
                          :  |  : | : | :        |:   ||  | |    | |     |
Db       1000 -------------MVFLKEQERQHLQDLLHQEVLRRIILLQRWFRVLLCRQHFLHLRQAS 1046

Qy        416 VKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDGL 475
              | : | |   |:     | |    |   |       |   | |: |   || |   :  :
Db       1047 VIIQRFWRNYLNQK----QVRDAAVQKDAFVMASAAALLQASWRAHLERQRYLELRAAAI 1102

Qy        476 --------------------DRWAVIAPQRNSHELR---TLLDSL---------YRAAS 502
                                   ||      :  | |   || |           ::|
Db       1103 VIQQKWRDYYRRRHMAAICIQARWKAYRESKRYQEQRKKIILLQSTCRGFRARQRFKALK 1162

Qy        503 GMGLRIRSPQEFII---------------------YDDRTGTYVRAMDDCVR------- 533
               ||   |:  ::                     :|:|    ::|:::|
Db       1163 EQRLRETKPEVGLVNIKGYGSLEIQGSDPSEWEDCSFDNR----IKAIEECKSVIESNRI 1218

Qy        534 SDPKLILCL--VPNDNAERYSS----------------------------IKKRGYVDR 562
              |    : ||   ||   ||  |                            |:     |
Db       1219 SRESSVDCLKESPNKQQERAQSQSGVDLQEDVLVRERPRSLEDLHQKKVGRAKRESRRMR 1278

Qy        563 AVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLGYTP-----WMIELPLSGLMTIGFDIAK 617
               :   : :|:  | |||               | :|     |  ||   || :      :
Db       1279 ELEQAIFSLELLKVRSLG-------------GISPSEDRRWSTELVPEGLQS-----PR 1319

Qy        618 STRDRKRAYGAL-IASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRKL 676
              | | : : |:| : | :  | |   | :::      :    | |: : :    ::
Db       1320 GTPDSESSQGSLELLSYEESQKSKLESVISDEGDLQFPS----PKISSSPKFDSRDNALS 1375

Qy        677 PSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLSPPQLAYIVVTRSMNTRFFL 736
              |      : :  |::|::     : |  |            |||    :: |: | |::
Db       1376 ASNETSSAEHLKDGTMKEMVVCSSESITCK-----------PQLKDSFISNSLPTFFYI 1423

Qy        737 NGQNP 741
                |:|
Db       1424 PQQDP 1428


RESULT 13
US-08-639-501-2
; Sequence 2, Application US/08639501
; Patent No. 5837492
;  GENERAL INFORMATION:
;    APPLICANT:  Tavtigian, Sean V.
```

```
;    APPLICANT:  Kamb, Alexander
;    APPLICANT:  Simard, Jacques
;    APPLICANT:  Couch, Fergus
;    APPLICANT:  Rommens, Johanna
;    APPLICANT:  Weber, Barbara
;    TITLE OF INVENTION:  Chromosome 13-Linked Breast Cancer
;    TITLE OF INVENTION:  Susceptibility Gene
;    NUMBER OF SEQUENCES:  124
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti
;      STREET:  1201 New York Avenue N.W., Suite 1001
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  22204
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/639,501
;      FILING DATE:  29-APR-1996
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/585,391
;      FILING DATE:  11-JAN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/576,559
;      FILING DATE:  21-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/575,359
;      FILING DATE:  20-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/573,779
;      FILING DATE:  18-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  24884-116802-04
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3418 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-639-501-2

  Query Match             2.3%;  Score 103;  DB 2;  Length 3418;
  Best Local Similarity   22.1%;  Pred. No. 6.8;
  Matches   85;  Conservative   44;  Mismatches  147;  Indels  108;  Gaps   17;

Qy         73 GDQYDYLNTRPVELVSKKGTDGVP----VMLQTNFFRXKTKPEWRIVHYHV--------- 119
              | |:::   |    : :| |  ||     :|:|      |  | |    ||
Db       1122 GSQFEFTQFRKPSYILQKSTFEVPENQMTILKT------TSEECRDADLHVIMNAPSIGQ 1175

Qy        120 -----EFVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKS---- 170
                  :|  |:|  |   |:| |  |   ||| |          | |:      |
Db       1176 VDSSKQFEGTVEIKRKFAGLLKNDCNKSASGYLTD----------ENEVGFRGFYSAHGT 1225

Qy        171 KLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFK 230
              ||::  :   | |   |  |       |:    | :   : |      |      :  ||
Db       1226 KLNVSTEALQKAVKLFSDIE-------NISEETSAEVHPISLSSSKCHDS----VVSMFK 1274

Qy        231 MELWPGYETS--------IRQHEKDILLGT---EITHKVMRT----ETIYDIMRRCSHN- 274
              :|       |        | |:  ::  ||   |||    |      :  |    | |||
Db       1275 IENHNDKTVSEKNNKCQLILQNNIEMTTGTFVEEITENYKRNTENEDNKYTAASRNSHNL 1334

Qy        275 -----------PARHQDEVRVNVLDLIVLTDYNNRTYR---------------INDVD 306
                         |:||      ||:  || :|   :                   ::|:

Db       1335 EFDGSDSSKNDTVCIHKDET-----DLL-FTDQHNICLKLSGQFMKEGNTQIKEDLSDLT 1388

Qy        307 FGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQ-----PLLISKNRDKALKTNASELV 361
              | :  |:  :| |   :  :   ||    |:|           ||   | | : :::|
Db       1389 FLEVAKAQEACHGNTSNKEQLTATKTEQNIKDFETSDTFFQTASGKNISVA-KESFNKIV 1447

Qy        362 VLI---PELCRVTGLNAEMRSNFQ 382
                    ||      ||:|: |: :
Db       1448 NFFDQKPEELHNFSLNSELHSDIR 1471


RESULT 14
US-09-044-946-2
; Sequence 2, Application US/09044946
; Patent No. 6033857
;  GENERAL INFORMATION:
;    APPLICANT:  Tavtigian, Sean V.
;    APPLICANT:  Kamb, Alexander
;    APPLICANT:  Simard, Jacques
;    APPLICANT:  Couch, Fergus
;    APPLICANT:  Rommens, Johanna
;    APPLICANT:  Weber, Barbara
;    TITLE OF INVENTION:  Chromosome 13-Linked Breast Cancer
;    TITLE OF INVENTION:  Susceptibility Gene
;    NUMBER OF SEQUENCES:  124
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti
;      STREET:  1201 New York Avenue N.W., Suite 1001
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  22204
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/044,946
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/639,501
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/576,559
;      FILING DATE:  21-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/575,359
;      FILING DATE:  20-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/573,779
;      FILING DATE:  18-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  24884-116802-04
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3418 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-044-946-2

  Query Match             2.3%;  Score 103;  DB 3;  Length 3418;
  Best Local Similarity   22.1%;  Pred. No. 6.8;
  Matches   85;  Conservative   44;  Mismatches  147;  Indels  108;  Gaps   17;

Qy         73 GDQYDYLNTRPVELVSKKGTDGVP----VMLQTNFFRXKTKPEWRIVHYHV--------- 119
```

```
           | |:::   |    : :| |  ||     :|:|      |  | |    ||
Db       1122 GSQFEFTQFRKPSYILQKSTFEVPENQMTILKT------TSEECRDADLHVIMNAPSIGQ 1175

Qy        120 -----EFVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKS---- 170
              :|  |:|  |   |:| |  |   |||| |          | |:      |
Db       1176 VDSSKQFEGTVEIKRKFAGLLKNDCNKSASGYLTD----------ENEVGFRGFYSAHGT 1225

Qy        171 KLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFK 230
              ||::  :   | |   |  |       |:    | :   : |      |       :  ||
Db       1226 KLNVSTEALQKAVKLFSDIE-------NISEETSAEVHPISLSSSKCHDS----VVSMFK 1274

Qy        231 MELWPGYETS--------IRQHEKDILLGT---EITHKVMRT----ETIYDIMRRCSHN- 274
              :|       |        | |:  ::  ||   |||    |     : |    | |||
Db       1275 IENHNDKTVSEKNNKCQLILQNNIEMTTGTFVEEITENYKRNTENEDNKYTAASRNSHNL 1334

Qy        275 ------------PARHQDEVRVNVLDLIVLTDYNNRTYR---------------INDVD 306
                          |:||      ||: || :|   :                  ::|:
Db       1335 EFDGSDSSKNDTVCIHKDET-----DLL-FTDQHNICLKLSGQFMKEGNTQIKEDLSDLT 1388

Qy        307 FGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQ-----PLLISKNRDKALKTNASELV 361
              | :  |:  :| |   :  :   ||    |:|             ||   | | : :::|
Db       1389 FLEVAKAQEACHGNTSNKEQLTATKTEQNIKDFETSDTFFQTASGKNISVA-KESFNKIV 1447

Qy        362 VLI---PELCRVTGLNAEMRSNFQ 382
                  ||      ||:|: |: :
Db       1448 NFFDQKPEELHNFSLNSELHSDIR 1471


RESULT 15
US-09-044-908-2
; Sequence 2, Application US/09044908
; Patent No. 6124104
;  GENERAL INFORMATION:
;    APPLICANT:  Tavtigian, Sean V.
;    APPLICANT:  Kamb, Alexander
;    APPLICANT:  Simard, Jacques
;    APPLICANT:  Couch, Fergus
;    APPLICANT:  Rommens, Johanna
;    APPLICANT:  Weber, Barbara
;    TITLE OF INVENTION:  Chromosome 13-Linked Breast Cancer
;    TITLE OF INVENTION:  Susceptibility Gene
;    NUMBER OF SEQUENCES:  124
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti
;      STREET:  1201 New York Avenue N.W., Suite 1001
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  22204
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/044,908
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/639,501
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/576,559
;      FILING DATE:  21-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/575,359
;      FILING DATE:  20-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/573,779
;      FILING DATE:  18-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  24884-116802-04
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3418 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-044-908-2

  Query Match             2.3%;  Score 103;  DB 3;  Length 3418;
  Best Local Similarity   22.1%;  Pred. No. 6.8;
  Matches   85;  Conservative   44;  Mismatches  147;  Indels  108;  Gaps   17;

Qy         73 GDQYDYLNTRPVELVSKKGTDGVP----VMLQTNFFRXKTKPEWRIVHYHV--------- 119
              | |:::   |    : :| |  ||     :|:|      |  | |    ||
Db       1122 GSQFEFTQFRKPSYILQKSTFEVPENQMTILKT------TSEECRDADLHVIMNAPSIGQ 1175

Qy        120 -----EFVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKS---- 170
              :|  |:|  |   |:| |  |   |||| |          | |:      |
Db       1176 VDSSKQFEGTVEIKRKFAGLLKNDCNKSASGYLTD----------ENEVGFRGFYSAHGT 1225

Qy        171 KLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFK 230
              ||::  :   | |   |  |       |:    | :   : |      |       :  ||
Db       1226 KLNVSTEALQKAVKLFSDIE-------NISEETSAEVHPISLSSSKCHDS----VVSMFK 1274

Qy        231 MELWPGYETS--------IRQHEKDILLGT---EITHKVMRT----ETIYDIMRRCSHN- 274
              :|       |        | |:  ::  ||   |||    |     : |    | |||
Db       1275 IENHNDKTVSEKNNKCQLILQNNIEMTTGTFVEEITENYKRNTENEDNKYTAASRNSHNL 1334

Qy        275 --------------PARHQDEVRVNVLDLIVLTDYNNRTYR---------------INDVD 306
                          |:||      ||: || :|   :                  ::|:
Db       1335 EFDGSDSSKNDTVCIHKDET-----DLL-FTDQHNICLKLSGQFMKEGNTQIKEDLSDLT 1388

Qy        307 FGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQ-----PLLISKNRDKALKTNASELV 361
              | :  |:  :| |   :  :   ||    |:|             ||   | | : :::|
Db       1389 FLEVAKAQEACHGNTSNKEQLTATKTEQNIKDFETSDTFFQTASGKNISVA-KESFNKIV 1447

Qy        362 VLI---PELCRVTGLNAEMRSNFQ 382
                  ||      ||:|: |: :
Db       1448 NFFDQKPEELHNFSLNSELHSDIR 1471


Search completed: July 22, 2003, 07:06:33
Job time : 14.1699 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 07:01:08 ; Search time 21.6828 Seconds
(without alignments)
4617.236 Million cell updates/sec

Title: US-09-873-737A-2
Perfect score: 4389
Sequence: 1 MADDQGRGRRRPXNEDDSST.....VGTNLHSIPQNALEKKFYYL 843

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep1:*
11: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep2:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep3:*
13: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4383	99.9	843	10	US-09-873-737A-2	Sequence 2, Appli
2	4349.5	99.1	844	14	US-10-043-774B-13	Sequence 13, Appl
3	4337.5	98.8	844	15	US-10-029-495-7	Sequence 7, Appli
4	1801.5	41.0	866	15	US-10-029-495-4	Sequence 4, Appli
5	1416.5	32.3	775	14	US-10-043-774B-2	Sequence 2, Appli
6	1406	32.0	861	10	US-09-873-737A-6	Sequence 6, Appli
7	1385.5	31.6	862	10	US-09-873-737A-4	Sequence 4, Appli
8	1087.5	24.8	824	15	US-10-029-495-6	Sequence 6, Appli
9	953.5	21.7	722	15	US-10-029-495-5	Sequence 5, Appli
10	931	21.2	523	11	US-09-998-598-2603	Sequence 2603, Ap
11	931	21.2	523	15	US-10-029-495-3	Sequence 3, Appli
12	875.5	19.9	580	10	US-09-801-574-16	Sequence 16, Appl
13	793.5	18.1	498	15	US-10-029-495-9	Sequence 9, Appli
14	472.5	10.8	915	15	US-10-174-363-40	Sequence 40, Appl
15	465.5	10.6	1101	15	US-10-174-363-54	Sequence 54, Appl
16	453.5	10.3	901	15	US-10-174-363-8	Sequence 8, Appli
17	444	10.1	904	15	US-10-174-363-55	Sequence 55, Appl
18	441	10.0	969	12	US-09-533-029-106	Sequence 106, App
19	441	10.0	969	15	US-10-286-264-102	Sequence 102, App
20	440.5	10.0	1065	15	US-10-174-363-28	Sequence 28, Appl
21	437	10.0	910	15	US-10-174-363-22	Sequence 22, Appl
22	431	9.8	1048	15	US-10-174-363-56	Sequence 56, Appl
23	431	9.8	1073	15	US-10-174-363-14	Sequence 14, Appl
24	426	9.7	1058	15	US-10-174-363-38	Sequence 38, Appl
25	425	9.7	1085	15	US-10-174-363-12	Sequence 12, Appl
26	421	9.6	1145	11	US-09-866-557A-5	Sequence 5, Appli
27	421	9.6	1145	15	US-10-055-797-5	Sequence 5, Appli
28	412.5	9.4	916	15	US-10-174-363-42	Sequence 42, Appl
29	410	9.3	821	15	US-10-174-363-4	Sequence 4, Appli
30	403.5	9.2	757	15	US-10-174-363-10	Sequence 10, Appl
31	366	8.3	919	15	US-10-174-363-16	Sequence 16, Appl
32	360	8.2	723	15	US-10-174-363-24	Sequence 24, Appl
33	356	8.1	577	15	US-10-174-363-2	Sequence 2, Appli
34	339.5	7.7	442	15	US-10-174-363-26	Sequence 26, Appl
35	311	7.1	389	15	US-10-174-363-50	Sequence 50, Appl
36	285.5	6.5	411	15	US-10-174-363-6	Sequence 6, Appli
37	245.5	5.6	186	15	US-10-029-495-2	Sequence 2, Appli
38	241.5	5.5	346	15	US-10-174-363-32	Sequence 32, Appl
39	168	3.8	130	15	US-10-102-806-685	Sequence 685, App
40	132	3.0	71	10	US-09-864-761-48051	Sequence 48051, A
41	113.5	2.6	645	10	US-09-731-231A-2	Sequence 2, Appli
42	112.5	2.6	630	15	US-10-128-714-3494	Sequence 3494, Ap
43	112	2.6	879	15	US-10-108-605-217	Sequence 217, App
44	111	2.5	55	10	US-09-864-761-38135	Sequence 38135, A
45	107.5	2.4	193	15	US-10-174-363-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1
US-09-873-737A-2
; Sequence 2, Application US/09873737A
; Patent No. US20020076797A1
; GENERAL INFORMATION:
; APPLICANT: Duke University
; APPLICANT: Lin, Haifan
; TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
; TITLE OF INVENTION: PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
; FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2
; CURRENT APPLICATION NUMBER: US/09/873,737A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/US99/28764
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/110,901
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 843
; TYPE: PRT
; ORGANISM: Drosophila sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (13)
; OTHER INFORMATION: Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (106)
; OTHER INFORMATION: Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (785)
; OTHER INFORMATION: Xaa=Leu or Ile
US-09-873-737A-2

Query Match 99.9%; Score 4383; DB 10; Length 843;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MADDQGRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAP 60

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MADDQGRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAP 60

Qy         61 RREGGPPERKPWGDQYDYLNTRPVELVSKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVE 120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RREGGPPERKPWGDQYDYLNTRPVELVSKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVE 120

Qy        121 FVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISI 180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 FVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISI 180

Qy        181 KFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETS 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETS 240

Qy        241 IRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTY 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTY 300

Qy        301 RINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASEL 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASEL 360

Qy        361 VVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLR 420
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLR 420

Qy        421 DWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDGLDRWAV 480
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDGLDRWAV 480

Qy        481 IAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCVRSDPKLIL 540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 IAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCVRSDPKLIL 540

Qy        541 CLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLGYTPWM 600
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLGYTPWM 600

Qy        601 IELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWP 660
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWP 660

Qy        661 MIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLSPPQ 720
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 MIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLSPPQ 720

Qy        721 LAYIVVTRSMNTRFFLNGQNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLY 780
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 LAYIVVTRSMNTRFFLNGQNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLY 780

Qy        781 SSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKF 840
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 SSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKF 840

Qy        841 YYL 843
                 |||
Db        841 YYL 843


RESULT 2
US-10-043-774B-13
; Sequence 13, Application US/10043774B
; Publication No. US20030129599A1
; GENERAL INFORMATION:
;  APPLICANT: University of Illinois at Chicago
;  APPLICANT:  Sharma, Arun
;  APPLICANT:  Hoffman, Ronald
;  TITLE OF INVENTION: HUMAN HEMATOPOIETIC GROWTH REGULATORY GENE AND USES
;  FILE REFERENCE: MBHB: CU08/PPA
;  CURRENT APPLICATION NUMBER: US/10/043,774B
;  CURRENT FILING DATE:  2002-06-06
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE:, PatentIn version 3.1
; SEQ ID NO 13
;   LENGTH: 844
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: PIWI protein
US-10-043-774B-13

  Query Match            99.1%;  Score 4349.5;  DB 14;  Length 844;
  Best Local Similarity  99.1%;  Pred. No. 0;
  Matches  836;  Conservative    1;  Mismatches    6;  Indels    1;  Gaps    1;

Qy          1 MADDQGRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAP 60
                 |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MADDQGRGRRRPLNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAP 60

Qy         61 RREGGPPERKPWGDQYDYLNTRPVELVSKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVE 120
                 |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db         61 RREGGPPERKPWGDQYDYLNTRPVELVSKKGTDGVPVMLQTNFFRLKTKPEWRIVHYHVE 120

Qy        121 FVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISI 180
                 | |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 FEPSIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISI 180

Qy        181 KFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETS 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETS 240

Qy        241 IRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTY 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTY 300

Qy        301 RINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASEL 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASEL 360

Qy        361 VVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLR 420
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLR 420

Qy        421 DWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDGLDRWAV 480
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDGLDRWAV 480

Qy        481 IAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCVRSDPKLIL 540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 IAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCVRSDPKLIL 540

Qy        541 CLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKN-RSLMSIATKIAIQLNCKLGYTPW 599
                 |||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db        541 CLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKKPYSLMSIATKIAIQLNCKLGYTPW 600

Qy        600 MIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLW 659
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 MIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLW 660

Qy        660 PMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLSPP 719
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLSPP 720

Qy        720 QLAYIVVTRSMNTRFFLNGQNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVL 779
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 QLAYIVVTRSMNTRFFLNGQNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVL 780

Qy        780 YSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKK 839
                 ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YSSMGLSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKK 840

Qy        840 FYYL 843
                 ||||
Db        841 FYYL 844
```

RESULT 3
US-10-029-495-7
; Sequence 7, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
; APPLICANT: Drmanac, Radoje T.
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Dickson, Mark C.
; APPLICANT: Jones, Lee William
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Xue, Aidong
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi,Vinod
; TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
; FILE REFERENCE: 28110/35905A
; CURRENT APPLICATION NUMBER: US/10/029,495
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 09/378,667
; PRIOR FILING DATE: 1999-08-20
; PRIOR APPLICATION NUMBER: 09/687,527
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 844
; TYPE: PRT
; ORGANISM: Drosophila
; FEATURE:
; OTHER INFORMATION: piwi
US-10-029-495-7

Query Match 98.8%; Score 4337.5; DB 15; Length 844;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 834; Conservative 1; Mismatches 8; Indels 1; Gaps 1;

```
Qy          1 MADDQGRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAP 60
              ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MADDQGRGRRRPLNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAP 60

Qy         61 RREGGPPERKPWGDQYDYLNTRPVELVSKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVE 120
              |||||| |||||||||||||||| |||||||||||||||||||| ||||||||||||||
Db         61 RREGGPTERKPWGDQYDYLNTRPAELVSKKGTDGVPVMLQTNFFRLKTKPEWRIVHYHVE 120

Qy        121 FVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISI 180
              | |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 FEPSIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISI 180

Qy        181 KFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETS 240

Qy        241 IRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTY 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTY 300

Qy        301 RINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASEL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASEL 360

Qy        361 VVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLR 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLR 420

Qy        421 DWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDGLDRWAV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDGLDRWAV 480

Qy        481 IAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCVRSDPKLIL 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 IAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCVRSDPKLIL 540

Qy        541 CLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKN-RSLMSIATKIAIQLNCKLGYTPW 599
              |||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db        541 CLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKKPYSLMSIATKIAIQLNCKLGYTPW 600

Qy        600 MIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLW 659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 MIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLW 660

Qy        660 PMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLSPP 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLSPP 720

Qy        720 QLAYIVVTRSMNTRFFLNGQNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVL 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 QLAYIVVTRSMNTRFFLNGQNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVL 780

Qy        780 YSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKK 839
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YSSMGLSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKK 840

Qy        840 FYYL 843
              ||||
Db        841 FYYL 844
```

RESULT 4
US-10-029-495-4
; Sequence 4, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
; APPLICANT: Drmanac, Radoje T.
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Dickson, Mark C.
; APPLICANT: Jones, Lee William
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Xue, Aidong
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi,Vinod
; TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
; FILE REFERENCE: 28110/35905A
; CURRENT APPLICATION NUMBER: US/10/029,495
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 09/378,667
; PRIOR FILING DATE: 1999-08-20
; PRIOR APPLICATION NUMBER: 09/687,527
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 866
; TYPE: PRT
; ORGANISM: Drosophila
; FEATURE:
; OTHER INFORMATION: aubergine
US-10-029-495-4

Query Match 41.0%; Score 1801.5; DB 15; Length 866;
Best Local Similarity 42.4%; Pred. No. 2.6e-166;
Matches 374; Conservative 159; Mismatches 279; Indels 71; Gaps 13;

```
Qy          5 QGRGR-RRPXN----------EDDSSTSRGS---------GD---GPRVK------V 32
              :|||| |:| |          : | | | |:        ||   || ::      |
Db         11 RGRGRGRKPNNVEANRGFAPSLGQKSDPSHSEGNQASGGNGGGGDAQVGPSIEKSSLSAV 70

Qy         33 FRGSSSGDPRADPRIEASRERRALEEAPRREGGPPERKPWGDQYDYLNTRPVELVSKKGT 92
                | ||||      : | || :                    | : :||   : ||||
Db         71 QMHKSEGDPRG-----SVRGRRLIT------------------DLVYSRPPGMTSKKGV 106

Qy         93 DGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMGVLSNHANLLGSGYLFDGLQ 152
```

```
Db        107 VGTHITVQANYFKVLKRPNWTIYQYRVDFTPDVEATRLRRSFLYEHKGILG-GYIFDGTN 165

Qy        153 LFTTRKFE-----QEITVLSGKSKLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKG 207

Db        166 MFCINQFKAVQDSPYVLELVTKSRAGENIEIKIKAVGSVQSTDAEQFQVLNLILRRAMEG 225

Qy        208 LNLELVGRNLFDPRAKIEIREFKMELWPGYETSIRQHEKDILLGTEITHKVMRTETIYDI 267

Db        226 LDLKLVSRYYYDPQAKINLENFRMQLWPGYQTSIRQHENDILLCSEICHKVMRTETLYNI 285

Qy        268 MRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTFSCKGRDISFVEY 327

Db        286 LSDAIRDSDDYQSTFKRAVMGMVILTDYNNKTYRIDDVDFQSTPLCKFKTNDGEISYVDY 345

Qy        328 YLTKYNIRIRDHNQPLLISKNRDKALKTNASELVVLIPELCRVTGLNAEMRSNFQLMRAM 387

Db        346 YKKRYNIIIRDLKQPLVMSRPTDKNIRGGNDQAIMIIPELARATGMTDAMRADFRTLRAM 405

Qy        388 SSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHN 447

Db        406 SEHTRLNPDRRIERLRMFNKRLKSCKQSVETLKSWNIELDSALVEIPARVLPPEKILFGN 465

Qy        448 GKV-PAGENADWQRHFRDQRMLTTPSDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGL 506

Db        466 QKIFVCDARADWTNEFRTCSMFKNVH--INRWYVITPSRNLRETQEFVQMCIRTASSMKM 523

Qy        507 RIRSPQEFIIYDDRTGTYVRAMDDCVRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPT 566

Db        524 NICNPIYEEIPDDRNGTYSQAIDNAAANDPQIVMVVMRSPNEEKYSCIKKRTCVDRPVPS 583

Qy        567 QVVTLKT-----TKNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRD 621

Db        584 QVVTLKVIAPRQQKPTGLMSIATKVVIQMNAKLMGAPWQVVIPLHGLMTVGFDVCHSPKN 643

Qy        622 RKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIV 681

Db        644 KNKSYGAFVATMDQKESFRYFSTVNEHIKGQELSEQMSVNMACALRSYQEQHRSLPERIL 703

Qy        682 FYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARV-QLSPPQLAYIVVTRSMNTRFFLNGQN 740

Db        704 FFRDGVGDGQLYQVVNSEVNTLKDRLDEIYKSAGKQEGCRMTFIIVSKRINSRYFTGHRN 763

Qy        741 PPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHL 800

Db        764 PVPGTVVDDVITLPERYDFFLVSQAVRIGTVSPTSYNVISDNMGLNADKLQMLSYKMTHM 823

Qy        801 YYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKFYYL 843

Db        824 YYNYSGTIRVPAVCHYAHKLAFLVAESINRAPSAGLQNQLYFL 866
```

RESULT 5
US-10-043-774B-2
; Sequence 2, Application US/10043774B
; Publication No. US20030129599A1
; GENERAL INFORMATION:
; APPLICANT: University of Illinois at Chicago
; APPLICANT: Sharma, Arun
; APPLICANT: Hoffman, Ronald
; TITLE OF INVENTION: HUMAN HEMATOPOIETIC GROWTH REGULATORY GENE AND USES
; FILE REFERENCE: MBHB: CU08/PPA
; CURRENT APPLICATION NUMBER: US/10/043,774B
; CURRENT FILING DATE: 2002-06-06
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 775
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-043-774B-2

Query Match 32.3%; Score 1416.5; DB 14; Length 775;

Best Local Similarity 37.7%; Pred. No. 8.2e-129;
Matches 292; Conservative 163; Mismatches 286; Indels 33; Gaps 14;

```
Qy         88 SKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMGVLSNHANLLGSGYL 147

Db         17 SKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHA 76

Qy        148 FDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKG 207

Db         77 FDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKI 136

Qy        208 LNLELVGRNLFDPRAKIEIREFKMELWPGYETSIRQHEKDILLGTEITHKVMRTETIYDI 267

Db        137 MNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDF 196

Qy        268 MRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTF-SCKGRDISF 324

Db        197 MFNFYHQTEEHKFQEQVSKELIGLVVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSF 256

Qy        325 VEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNASELVVLIPELCRVTGLNAEMRSNFQL 383

Db        257 LEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNV 316

Qy        384 MRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQ 440

Db        317 MKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQ---RELRDWGLSFDSNLLSFSGRILQT 373

Qy        441 QNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTPSDGLDRWAVIAPQRNSHELRTLLDSL 497

Db        374 EKI--HQGGKTFDYNPQFADWSKETRGAPLISVKP--LDNWLLIYTRRNYEAANSLIQNL 429

Qy        498 YRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCVRSDPKLILCLVPNDNAERYSSIKKR 557

Db        430 FKVTPAMGMQMRK-AIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKY 488

Qy        558 GYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIA 616

Db        489 LCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCY 548

Qy        617 KSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRKL 676

Db        549 HDMTAGRRSIAGFVASIN-EGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYM 607

Qy        677 PSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKT---EYARVQLSPPQLAYIVVTRSMNTR 733

Db        608 PSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGY------NPRLTVIVVKKRVNTR 661

Qy        734 FFLNG----QNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEK 789

Db        662 FFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDH 721

Qy        790 MQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKFYYL 843

Db        722 IQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 775
```

RESULT 6
US-09-873-737A-6
; Sequence 6, Application US/09873737A
; Patent No. US20020076797A1
; GENERAL INFORMATION:
; APPLICANT: Duke University
; APPLICANT: Lin, Haifan
; TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
; TITLE OF INVENTION: PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
; FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2
; CURRENT APPLICATION NUMBER: US/09/873,737A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/US99/28764
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/110,901
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 21

```
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 861
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (76)
;   OTHER INFORMATION: Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (303)
;   OTHER INFORMATION: Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (735)
;   OTHER INFORMATION: Xaa=Leu or Ile
US-09-873-737A-6

  Query Match             32.0%;  Score 1406;  DB 10;  Length 861;
  Best Local Similarity   35.8%;  Pred. No. 1e-127;
  Matches  308;  Conservative  172;  Mismatches  312;  Indels   68;  Gaps   20;

Qy          6 GRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG 65
              ||||:|       |: |:     :::  |                          :|    | |
Db         48 GRGRQR-------GTAGGTAKSQGLQISAG-------------------FQELSXAERG 80

Qy         66 PPERKPWGDQYDY-LNTRP-VELV--SKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEF 121
                 |    | :|  :|||  :: |  || |: |: | | || ||  ::|:| :  ||:::
Db         81 GRRR----DFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDY 136

Qy        122 VPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIK 181
               | :|  |:|  :|  | :|:|  : |||  ||  :: :|::| :  |::   : :|:|
Db        137 NPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITIT 196

Qy        182 FVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSI 241
                  :    |  ||  |:| || :| :||: :||| ::|   |:|   :: :|||: |||
Db        197 LTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSI 256

Qy        242 RQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRT 299
               |:|  |:| |:::|||:|:||: | |     |     |  |::|    : |:|||  || :|
Db        257 LQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKT 316

Qy        300 YRINDVDFGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNA 357
              ||::|:|: | |||||    |  :||:|||  :||  | |  ||:|:|:  | :
Db        317 YRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTL 376

Qy        358 SELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPE 414
                 :|||||| :|||  :||::| :|: :: :||: |:||    ||  : |:  |
Db        377 PGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQ- 435

Qy        415 SVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTP 471
               : |||| :  | |:    |||:  : |  | |       |   |||  :  |   :::
Db        436 --RELRDWGLSFDSNLLSFSGRILQTEKI--HQGGKTFDYNPQFADWSKETRGAPLISVK 491

Qy        472 SDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDC 531
               || | :|  :||     :|: :|::    ||:::|      |  ||||  |:| :
Db        492 P--LDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRK-AIMIEVDDRTEAYLRVLQQK 548

Qy        532 VRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQL 590
               | :| ::::||: ::  ::| :|||     |   |:| |  :|  | :::|:|||||:|:
Db        549 VTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQM 608

Qy        591 NCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSA 650
              |||:|   | :::||  :| :| |        :|:    :||:: :  : :||
Db        609 NCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASIN-EGMTRWFSRCIFQDR 667

Qy        651 FDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKT- 709
               | : |   :  ||| :   :  :|||| |||||  | || |  :||   :: ||:
Db        668 GQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSI 727

Qy        710 --EYARVQLSPPQLAYIVVTRSMNTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVS 763
                   |      |:   ||| :  :|||||       ||| |||::|  :| || |||::||
Db        728 GRGY------NPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVS 781

Qy        764 QQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATL 823
               | || |:|||| |||:| : |  |: :|:|||||:||:|||| |  |||| |||| ||| |
Db        782 QAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFL 841

Qy        824 VGTNLHSIPQNALEKKFYYL 843
               || ::|  |  :|  : |||
Db        842 VGQSIHREPNLSLSNRLYYL 861


RESULT 7
US-09-873-737A-4
; Sequence 4, Application US/09873737A
; Patent No. US20020076797A1
; GENERAL INFORMATION:
;  APPLICANT: Duke University
;  APPLICANT:  Lin, Haifan
;  TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
;  TITLE OF INVENTION:  PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
;  FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2
;  CURRENT APPLICATION NUMBER: US/09/873,737A
;  CURRENT FILING DATE:  2001-06-04
;  PRIOR APPLICATION NUMBER: PCT/US99/28764
;  PRIOR FILING DATE: 1999-12-03
;  PRIOR APPLICATION NUMBER: 60/110,901
;  PRIOR FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 862
;   TYPE: PRT
;   ORGANISM: Mus sp.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (90)
;   OTHER INFORMATION: Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (216)
;   OTHER INFORMATION: Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (383)
;   OTHER INFORMATION: Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (816)
;   OTHER INFORMATION: Xaa=Leu or Ile
US-09-873-737A-4

  Query Match             31.6%;  Score 1385.5;  DB 10;  Length 862;
  Best Local Similarity   35.7%;  Pred. No. 1e-125;
  Matches  306;  Conservative  173;  Mismatches  318;  Indels   59;  Gaps   19;

Qy          6 GRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG 65
              ||||:|       :| |:                          : :| |  :  :| |     ||
Db         48 GRGRQRGMVVGATSKSQ--------------------ELQISAGFQELSLAE----RGG 82

Qy         66 PPERKPWGDQYDYLNTRP-VELV--SKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFV 122
                 |: : |     :|||  :: |  || |: |: | | || ||  ::|:| :  ||:::
Db         83 --RRRDFHD--XGVNTRQNLDHVKESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYN 138

Qy        123 PTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIKF 182
               | :|  |:|  :|  | :|:|  : |||  ||  :: : ::| :  :::      :|:|
Db        139 PLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVTEVFSQTRNGEHVRITITL 198

Qy        183 VGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSIR 242
                  :    |  ||  |:  || :| :||: :||| ::|   |:|   :: :|||: |||
Db        199 TNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSIL 258

Qy        243 QHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRTY 300
               |:| :|:| |:::|||:|:||: | |          |  |::|   :: |||||  |||:||
Db        259 QYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTY 318

Qy        301 RINDVDFGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNAS 358
              |::|:|: | |||||    | ::||:|||  :||  | |  ||:|:|:  | :
Db        319 RVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLP 378
```

```
Qy        359 ELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPES 415
Db        379 GPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQ-- 436

Qy        416 VKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTPS 472
Db        437 -RELRDWGLSFDSNLLSFSGRILQSEKI--HQGGKTFDYNPQFADWSKETRGAPLISVKP 493

Qy        473 DGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCV 532
Db        494 --LDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKK-AIMIEVDDRTEAYLRALQQKV 550

Qy        533 RSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQLN 591
Db        551 TSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMN 610

Qy        592 CKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAF 651
Db        611 CKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASIN-EGMTRWFSRCVFQDRG 669

Qy        652 DVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEY 711
Db        670 QELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVG 729

Qy        712 ARVQLSPPQLAYIVVTRSMNTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVSQQVR 767
Db        730 RGYN---PRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVR 786

Qy        768 QGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTN 827
Db        787 SGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVGQS 846

Qy        828 LHSIPQNALEKKFYYL 843
Db        847 IHREPNLSLSNRLYYL 862
```

RESULT 8
US-10-029-495-6
; Sequence 6, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
; APPLICANT: Drmanac, Radoje T.
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Dickson, Mark C.
; APPLICANT: Jones, Lee William
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Xue, Aidong
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi,Vinod
; TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
; FILE REFERENCE: 28110/35905A
; CURRENT APPLICATION NUMBER: US/10/029,495
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 09/378,667
; PRIOR FILING DATE: 1999-08-20
; PRIOR APPLICATION NUMBER: 09/687,527
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 824
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:
; OTHER INFORMATION: ciwi2
US-10-029-495-6

Query Match 24.8%; Score 1087.5; DB 15; Length 824;
Best Local Similarity 29.9%; Pred. No. 1.2e-96;

Matches 267; Conservative 167; Mismatches 341; Indels 117; Gaps 22;

```
Qy          1 MADDQGRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAP 60
Db          1 MASGSGRGR-----------GRGSG---------SNNSG---------------------- 19

Qy         61 RREGGPPERKPWGDQYDYLNTRPVELV------SKKGTDGVPVMLQTNFFRXK-TKPEWR 113
Db         20 ------------GKDQKYLGTIQPDLFIRQQGQSKTGSSGQPQKCFANFIPIEMTQSDYS 67

Qy        114 IVHYHVEFVPTIENPRVRMGVL--SNHANLLGSGYLFDGLQLFTTRKFEQ----EITVLS 167
Db         68 IYQYHVEFEPTVDSKANREKMLRDNNVTDEIGHHFVFDGMILYLKEEWEQNQMIEVQHPI 127

Qy        168 GKSKLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLF---DPRAKI 224
Db        128 DRSLICIRFKQTNRFL----VDDPQTINIFNTIIRRSFDALQLTQLGRNYFNWGDSRA-- 181

Qy        225 EIREFKMELWPGYETSIRQHEKDILLGTEITHKVMRTETIYDI----MRRCSHNPARHQD 280
Db        182 -VPDYNMSILPGYETAIRMYEENFMLCVENRFKMVREESMYILFHKELRSCQNNPQRVQE 240

Qy        281 EVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHN 340
Db        241 KMNEMYGGTTIITRYNNKLHRYTRLDYSISPLSEFVKDGQSITLKEYFKNQYGIEITVDD 300

Qy        341 QPLLISKNRDKALKTNASELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTD 400
Db        301 QPIIISEGKPKQ-PGEPPQVSYIVPELCFPTGLTDEMRKDFKMMKEIAKHTRMSPQQRLV 359

Qy        401 RLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQR 460
Db        360 ESRKLIVDLSKNEKVMECFKYWGISLGQDLANVQARVLKSEPL--QGKKTYEGKQAEWAR 417

Qy        461 HFRDQRMLTTPSDGLDRWAVIAPQRNSHEL--RTLLDSLYRAASGMGLRIRSPQEFIIYD 518
Db        418 GVKECGIYR--GSNMTNWIVIGPGSGNSGLLSQKFIEEARRLGKILQVQLGEPMCVPIRG 475

Qy        519 DRTGTYVRAMDDCVR----SDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT 574
Db        476 ISPNDYLEGVKGAIKQVDGEDIHMLVVMLADDNKTRYDSLKKFLCVECPIPNQCVNLRTL 535

Qy        575 --------KNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAY 626
Db        536 AGKSKDGGENKNLGSIVLKIVLQMICKTGGALWKVNIPLKNTMIVGYDLYHDSTLKGKTV 595

Qy        627 GALIA--SMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEH-RKLPSRIVFY 683
Db        596 GACVSTTSNDFTQ---FYSQTRPHENPTQLGNNLTHFVRKALKQYYDSNDQTLPSRLILY 652

Qy        684 RDGVSSGSLKQLFEFEVK------DIIEKLKTEYARVQLSPPQLAYIVVTRSMNTRFFLN 737
Db        653 RDGAGDGQIPYIKNTEVKLVRDACDAVTDKAAELSNKVQEKIKLAFIIVTKRVNMRILKQ 712

Qy        738 GQ------NPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQ 791
Db        713 GSSLDNAINPQPGTVVDTTVTRPERMDFYLVPQFVNQGTVTPVSYNIIHDDTDLGPDKHQ 772

Qy        792 KLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKFYYL 843
Db        773 QLAFKLCHLYYNWQGTVRVPAPCQYAHKLAFLTAQSLHDDANGCLRDKLFFL 824
```

RESULT 9
US-10-029-495-5
; Sequence 5, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
; APPLICANT: Drmanac, Radoje T.
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Dickson, Mark C.
; APPLICANT: Jones, Lee William

```
;  APPLICANT:  Ballinger, Dennis G.
;  APPLICANT:  Xue, Aidong
;  APPLICANT:  Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi,Vinod
;  TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
;  FILE REFERENCE: 28110/35905A
;  CURRENT APPLICATION NUMBER: US/10/029,495
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: 09/378,667
;  PRIOR FILING DATE: 1999-08-20
;  PRIOR APPLICATION NUMBER: 09/687,527
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 722
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
;   FEATURE:
;   OTHER INFORMATION: ciwi1
US-10-029-495-5

  Query Match             21.7%;  Score 953.5;  DB 15;  Length 722;
  Best Local Similarity   29.8%;  Pred. No. 1.1e-83;
  Matches  230;  Conservative  144;  Mismatches  306;  Indels   91;  Gaps   19;

Qy        111 EWRIVHYHVEFVPTIENPRVRMGVLSNHANL--LGSGYLFDGLQLFTTRKFEQ----EIT 164
              :: |  ||||| ||:::   |  :|   :    :|  ::|||: |:   :::|    |:
Db          5 DYSIYQYHVEFEPTVDSKATRENMLRQPSVTVEIGKHFVFDGMILYLKEEWDQNQMIEVQ 64

Qy        165 VLSGKSKLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKI 224
                :  | : | :| : :|:      :|: : : | |:|||   : |  :||| ||
Db         65 HPNDNSLICIRFKKTNRFL----VDDPQTINIFNTIIRRSFDAMKLTQIGRNYFD----- 115

Qy        225 EIREFKMELWPGYETSIRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARHQDEVRV 284
                  |                    : :| |     :| | :|  | |:::
Db        116 ---------WD------------------NSRALRKE-----LRSCQNNRQRVQEKMNE 142

Qy        285 NVLDLIVLTDYNNRTYRINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLL 344
                ::| |||: :|   :|   || | |   |  |   ||:  :|:| | |  | ::
Db        143 VYGGSTIITRYNNKLHRFTRLDNEITPLSKFQKDGEQIILKEYFKNQYDIDITDDEQFII 202

Qy        345 ISKNRDKALKTNASELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRA 404
              ||: : |       ::  ::|||| |||  ||| :|:::|: :: :|||:|:|| |  |
Db        203 ISEGKPKQ-PGEPPQVNYIVPELCFPTGLTDEMRKDFKMMKEIAKHTRMSPQQRLDETRK 261

Qy        405 FNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVP-AGENADWQRHFR 463
                :|      ::  : | : | ::|  || |::  : :    ||    |: |:| | :
Db        262 LITKLSQNQTMMECFQYWGISLGQDLANVQARVLKSEPL---QGKRQYEGKQAEWARGVK 318

Qy        464 DQRMLTTPSDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMG----LRIRSPQEFIIYDD 519
              :  :       :  | || |   :  | |       | :|    :::  |    |
Db        319 ECGIYR--GSNMTNWIVIGPGSGNSGL--LAQKFIAEARNLGRTLQVQLGEPMCVKINGI 374

Qy        520 RTGTYVRAMDDCVRS----DPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT- 574
                |: :   ::|    :  ::: :: :||  || |:||   |:  :| | | |:|
Db        375 SPNDYLEGLKAAIKSVDGEEIHMLVVMLADDNKTRYDSLKKYLCVECPIPNQCVNLRTLA 434

Qy        575 -------KNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYG 627
                   :|:: | ||  || :|: || |   | : :||   | :|:|:   :  : :  |
Db        435 GKSKDGGENKNLGSIVLKIVLQMICKTGGALWKVNIPLKSTMIVGYDLYHDSTLKGKTVG 494

Qy        628 ALIA--SMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRK-LPSRIVFYR 684
              | ::  | | |   ::|          | | |   : |:|:||   : | ||||:: ||
Db        495 ACVSTTSNDFTQ---FYSQTRPHENPTQLGNNLTHFVRKSLKQYYDNNDKTLPSRLILYR 551

Qy        685 DGVSSGSLKQLFEFEVK------DIIEKLKTEYARVQLSPPQLAYIVVTRSMNTRFFLNG 738
              ||   | :  :   |||       | :      | :        :|||:|:||: :| |    |
Db        552 DGAGDGQIPYIKNTEVKLVRDACDAVTDKAAELSNKVQEKIKLAFIIVTKRVNMRILKQG 611

Qy        739 Q------NPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQK 792
                   || |||:||  :| ||| ||||| | | ||||:| |||:::   |  |:| |:

Db        612 SSSKSAINPQPGTVVDTTVTRPERMDFYLVPQFVNQGTVTPVSYNIIHDDTGLGPDKHQQ 671

Qy        793 LTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKFYYL 843
              | :|:|||||| || |||| |||| ||| |   :||       |  | ::|
Db        672 LAFKLCHLYYNWQGTVRVPAPCQYAHKLAFLTAQSLHDDANGYLRDKLFFL 722


RESULT 10
US-09-998-598-2603
; Sequence 2603, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
;  APPLICANT: Stolk, John A.
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Chenault, Ruth A.
;  APPLICANT:  Meagher, Madelein Joy
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.561
;  CURRENT APPLICATION NUMBER: US/09/998,598
;  CURRENT FILING DATE:  2001-11-16
;  NUMBER OF SEQ ID NOS: 2606
;  SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 2603
;   LENGTH: 523
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-998-598-2603

  Query Match             21.2%;  Score 931;  DB 11;  Length 523;
  Best Local Similarity   37.6%;  Pred. No. 1e-81;
  Matches  202;  Conservative  105;  Mismatches  200;  Indels   30;  Gaps   12;

Qy        322 ISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNASELVVLIPELCRVTGLNAEMRSN 380
              :||:|||  :||  |'|  ||:|:|:  | :           :|||||| :|||  :||::
Db          2 VSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRND 61

Qy        381 FQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRI 437
              . | :|: :: :|||: |:||     ||  : |:  |     : |||| :  | |:    |||
Db         62 FNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQ---RELRDWGLSFDSNLLSFSGRI 118

Qy        438 IGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTPSDGLDRWAVIAPQRNSHELRTLL 494
              :  : |  | |       |   |||  :  |   :::     || | :|  :||      :|:
Db        119 LQTEKI--HQGGKTFDYNPQFADWSKETRGAPLISVKP--LDNWLLIYTRRNYEAANSLI 174

Qy        495 DSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCVRSDPKLILCLVPNDNAERYSSI 554
              :|::    ||:::|      |  ||||  |:| :   | :| ::::||: ::  ::| :|
Db        175 QNLFKVTPAMGMQMRK-AIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAI 233

Qy        555 KKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGF 613
              ||     |  |:| |  :|  | :::|:|||||:|:|||:|   | :::||  :| :|
Db        234 KKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGI 293

Qy        614 DIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEH 673
              |        :|:    :||:: :  : :||           | : |   :  ||| :  :
Db        294 DCYHDMTAGRRSIAGFVASIN-EGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCN 352

Qy        674 RKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKT---EYARVQLSPPQLAYIVVTRSM 730
                :|||: |||||  | || |  :||    :: ||:    |        |:|  ||| : :
Db        353 EYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGY------NPRLTVIVVKKRV 406

Qy        731 NTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXS 786
              |||||       ||| |||::|  :| || |||::||| || |:|||| |||:| : |
Db        407 NTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLK 466

Qy        787 PEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKFYYL 843
              |: :|:|||||:||:|||| |  |||| |||| ||| ||| ::|  |  :|  : |||
Db        467 PDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 523


RESULT 11
US-10-029-495-3
; Sequence 3, Application US/10029495
```

```
; Publication No. US20030027255A1
; GENERAL INFORMATION:
;  APPLICANT: Drmanac, Radoje T.
;  APPLICANT:  Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Dickson, Mark C.
;  APPLICANT:  Jones, Lee William
;  APPLICANT:  Ballinger, Dennis G.
;  APPLICANT:  Xue, Aidong
;  APPLICANT:  Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi,Vinod
;  TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
;  FILE REFERENCE: 28110/35905A
;  CURRENT APPLICATION NUMBER: US/10/029,495
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: 09/378,667
;  PRIOR FILING DATE: 1999-08-20
;  PRIOR APPLICATION NUMBER: 09/687,527
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 523
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: hiwi
US-10-029-495-3

  Query Match           21.2%;  Score 931;  DB 15;  Length 523;
  Best Local Similarity  37.6%;  Pred. No. 1e-81;
  Matches  202;  Conservative  105;  Mismatches  200;  Indels   30;  Gaps   12;

Qy        322 ISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNASELVVLIPELCRVTGLNAEMRSN 380
              :||:|||  :||  | |  ||:|:|:  | :          :|||||| :|||  :||::
Db          2 VSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRND 61

Qy        381 FQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRI 437
              | :|: :: :||: |:||     ||  : |:  |     : |||| :  | |:    |||
Db         62 FNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQ---RELRDWGLSFDSNLLSFSGRI 118

Qy        438 IGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTPSDGLDRWAVIAPQRNSHELRTLL 494
              :  :  | |       |   ||| :  |   :::     || | :|  :||     :|:
Db        119 LQTEKI--HQGGKTFDYNPQFADWSKETRGAPLISVKP--LDNWLLIYTRRNYEAANSLI 174

Qy        495 DSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCVRSDPKLILCLVPNDNAERYSSI 554
              :|::     ||:::|     |  ||||  |:| :   | :| ::::||: ::  ::| :|
Db        175 QNLFKVTPAMGMQMRK-AIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAI 233

Qy        555 KKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGF 613
              ||    |   |:| |  :|  | :::|:|||||||:|:|||:|   | :::||  :| :|
Db        234 KKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGI 293

Qy        614 DIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEH 673
              |        :|:    :||:: :  : :||          | : |   :  |||  :   :
Db        294 DCYHDMTAGRRSIAGFVASIN-EGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCN 352

Qy        674 RKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKT---EYARVQLSPPQLAYIVVTRSM 730
              :||||: |||||  | || |  :||   :: ||:     |        |:|  ||| : :
Db        353 EYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGY------NPRLTVIVVKKRV 406

Qy        731 NTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXS 786
              |||||       ||| |||::|  :| || |||::||| || |:|||| |||:| : |
Db        407 NTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLK 466

Qy        787 PEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKFYYL 843
              |: :|:||||:||:|||| |  |||| |||| ||| ||| ::|  |  :|  : |||
Db        467 PDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 523


RESULT 12
US-09-801-574-16
; Sequence 16, Application US/09801574
; Patent No. US20020081592A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Peijing Jeremy
;  APPLICANT:  Page, David C.
;  TITLE OF INVENTION: Reproduction-Specific Genes
;  FILE REFERENCE: 0399.2007-002
;  CURRENT APPLICATION NUMBER: US/09/801,574
;  CURRENT FILING DATE:  2001-03-07
;  PRIOR APPLICATION NUMBER: 60/187,518
;  PRIOR FILING DATE: 2000-03-07
;  PRIOR APPLICATION NUMBER: 60/261,557
;  PRIOR FILING DATE: 2001-01-12
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
;   LENGTH: 580
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-801-574-16

  Query Match           19.9%;  Score 875.5;  DB 10;  Length 580;
  Best Local Similarity  33.8%;  Pred. No. 3.3e-76;
  Matches  197;  Conservative  120;  Mismatches  243;  Indels   23;  Gaps   10;

Qy        274 NPARHQDEVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTF-SCKGRDISFVEYYLTKY 332
              |   |||    ::  ||:| |||||||:|||: :||| :|    |::|:|:|||   |
Db          8 NKEHFQDECSKLLVGSIVITRYNNRTYRIDDVDWNKTPKDSFVMSDGKEITFLEYYSKNY 67

Qy        333 NIRIRDHNQPLLISKNRDKALKTNASEL----VVLIPELCRVTGLNAEMRSNFQLMRAMS 388
                | ::: :|||||  :|    : |   |    ::|:|||  :||:  :|: :|: |: ::
Db         68 GITVKEDDQPLLI--HRPSERQNNHGMLLKGEILLLPELSFMTGIPEKMKKDFRAMKDLT 125

Qy        389 SYTRMNPKQRTDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNG 448
              ::|||    |     |:     :   |  | : | |:| :::||::  : |   |
Db        126 QQINLSPKQHHGALECLLQRISQNETASNELTRWGLSLHKDVHKIEGRLLPMERINLRNT 185

Qy        449 KVPAGENADWQRHF-RDQRMLTTPSDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLR 507
                |  :| :   ||  :|| |   :  ||:  |:|   : | |:: | : |  :|:|
Db        186 SFVTSEGLNWVKEVTRDASILTIP---MHFWALFYPKRAMDQARELVNMLEKIAGPIGMR 242

Qy        508 IRSPQEFIIYDDRTGTYVRAMDDCVRSDPK--LILCLVPNDNAERYSSIKKRGYVDRAVP 565
              |    : |||  ||:| :   :  : |  :::|::     : | :|||   |   ||
Db        243 TSPPAWVELKDDRIETYIRTIQSLLGVEGKIQMVVCIIMGTRDDLYGAIKKLCCVQSPVP 302

Qy        566 TQVVTLKTT-KNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKR 624
              :|: ::|  :   | |:| || :|:|||||   | :::||  || || |:          |
Db        303 SQVINVRTIGQPTRLRSVAQKILLQMNCKLGGELWGVDIPLKQLMVIGMDVYHDPSRGMR 362

Qy        625 AYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYR 684
              :    :||::|     |   | :   ::: ::|   :  :|::|   :  || :|| ||
Db        363 SVVGFVASINLTLTKWYSRVVFQMPHQEIV-DSLKLCLVGSLKKYYEVNHCLPEKIVVYR 421

Qy        685 DGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLSPPQLAYIVVTRSMNTRFFLNGQN---- 740
              |||| | || :  :|    | :|:  :       |::   || : ::|  :|   :
Db        422 DGVSDGQLKTVANYE----IPQLQKCFEAFDNYHPKMVVFVVQKKISTNLYLAAPDHFVT 477

Qy        741 PPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHL 800
              | |||:||  ||  |  ||||::  ||||   || |  : ::   ||: ||:||:|:||:
Db        478 PSPGTVVDHTITSCEWVDFYLLAHHVRQGCGIPTHYICVLNTANLSPDHMQRLTFKLCHM 537

Qy        801 YYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKFYYL 843
              |:|| || |||| |:|| ||| | |  ||  |   |    ::|
Db        538 YWNWPGTIRVPAPCKYAHKLAFLSGQILHHEPAIQLCGNLFFL 580


RESULT 13
US-10-029-495-9
; Sequence 9, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
;  APPLICANT: Drmanac, Radoje T.
;  APPLICANT:  Labat, Ivan
```

; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Dickson, Mark C.
; APPLICANT: Jones, Lee William
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Xue, Aidong
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi,Vinod
; TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
; FILE REFERENCE: 28110/35905A
; CURRENT APPLICATION NUMBER: US/10/029,495
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 09/378,667
; PRIOR FILING DATE: 1999-08-20
; PRIOR APPLICATION NUMBER: 09/687,527
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-029-495-9

Query Match 18.1%; Score 793.5; DB 15; Length 498;
Best Local Similarity 35.2%; Pred. No. 2.5e-68;
Matches 178; Conservative 89; Mismatches 226; Indels 13; Gaps 7;

```
Qy        343 LLISKNRDKALKTNASELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRL 402
              :|:|  : |    :  :|  |||||| :|||  :  |:||||:|::  ||::|  |  ||
Db          1 MLVSLLKKKRNDNSEPQLAHLIPELCFLTGLTDQATSDFQLMKAVAEKTRLSPSGRQQRL 60

Qy        403 RAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHF 462
                    :|    :   |  | :     :: : |||:  : |:  :         ||| :
Db         61 ARLVDNIQRNTNARFELETWGLHFGSQIS-LTGRIVPSEKILMQDHICQPVSAADWSKDI 119

Qy        463 RDQRMLTTPSDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTG 522
              |  ::|   |  |: | ::   |  :   : |: | | |  ||  :  | : |   :
Db        120 RTCKILNAQS--LNTWLILCSDRTEYVAESFLNCLRRVAGSMGFNVDYP-KIIKVQENPA 176

Qy        523 TYVRAMDDCVRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKT-TKNRSLMS 581
               :|||:   |  | :|::|::|::    | ||||    |  ||:| |  :|  |   :||
Db        177 AFVRAIQQYVDPDVQLVMCILPSNQKTYYDSIKKYLSSDCPVPSQCVLARTLNKQGMMMS 236

Qy        582 IATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTY 641
              ||||||:|: ||||   | :|:||  || :| |: |    :        :||:: :    :
Db        237 IATKIAMQMTCKLGGELWAVEIPLKSLMVVGIDVCKDALSKDVMVVGCVASVNPRITRWF 296

Qy        642 FSTVTECSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVK 701
                : :  || |: |   :  || ::   :  ||:||: || ||  | || | |:||
Db        297 SRCILQRTMTDV-ADCLKVFMTGALNKWYKYNHDLPARIIVYRAGVGDGQLKTLIEYEVP 355

Qy        702 DIIEKLKTEYARVQLSPPQLAYIVVTRSMNTRFFL----NGQNPPPGTIVDDVITLPERY 757
               ::  :    :    :  :|: ||| :   |||       |||| ||:||   |  | |
Db        356 QLLSSVAESSSN---TSSRLSVIVVRKKCMPRFFTEMNRTVQNPPLGTVVDSEATRNEWY 412

Qy        758 DFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYA 817
              ||||:||   :|||||| |||:|   |  |: ||:||:|:|||||| |   ||| ||||
Db        413 DFYLISQVACRGTVSPTYYNVIYDDNGLKPDHMQRLTFKLCHLYYNWPGIVSVPAPCQYA 472

Qy        818 KKLATLVGTNLHSIPQNALEKKFYYL 843
               ||  |   ::|  |   |    :||
Db        473 HKLTFLGAQSIHKEPSLELANHLFYL 498
```

RESULT 14
US-10-174-363-40
; Sequence 40, Application US/10174363
; Publication No. US20030077623A1
; GENERAL INFORMATION:
; APPLICANT: Butler, Karlene H.
; APPLICANT: Cahoon, Rebecca E
; APPLICANT: Harvell, Leslie T.
; APPLICANT: Rafalski, Antoni J.
; APPLICANT: Sakai, Hajime
; TITLE OF INVENTION: Polynucleotides And Polypeptides Involved In Post-Transcriptio
; TITLE OF INVENTION: Gene Silencing
; FILE REFERENCE: BB1454 US NA
; CURRENT APPLICATION NUMBER: US/10/174,363
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: 60/298,973
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 40
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Glycine max
US-10-174-363-40

Query Match 10.8%; Score 472.5; DB 15; Length 915;
Best Local Similarity 22.7%; Pred. No. 1.6e-36;
Matches 197; Conservative 156; Mismatches 359; Indels 155; Gaps 32;

```
Qy         83 PVELVSKK----------GTDGVPVMLQTNFFRXK-TKPEWRIVHYHVEFVPTIENPRV 130
               | |   ||           |: |  : | || |:    | :    || | |      |
Db         37 PTEHNKKKASRLPIARSGLGSKGNKIQLLTNHFKVNVAKNDGHFFHYSVAFTYEDGRPVE 96

Qy        131 RMGV---------LSNHANLLGSGYLFDG-LQLFTT-----RKFEQEIT---VLSGKS-- 170
                ||          : |::| |  : :||   |||       | | |:    | | ::
Db         97 GKGVGRKIIDRVQETYHSDLNGKDFAYDGEKSLFTVGSLPQNKLEFEVVLEDVTSNRNNG 156

Qy        171 -------------------------KLDIEYKISIKFVGFISCAE-------PRFLQVL 197
                                       |::| :   |    |                ::||
Db        157 NCSPDGLGDNESDRKRMRRPYRSKSFKVEISFAAKIPMQAIASALRGQETENFQEAIRVL 216

Qy        198 NLILRRSMKGLNLELVGRNLF--DPRAKIEIREFKMELWPGYETSIRQHEKDILLGTEI- 254
              ::|||:        || :: |  :|    ::    :    |: :| |  :  : |  ::
Db        217 DIILRQHAAKQGCLLVRQSFFHNNPNNFADVGGGVLGC-RGFHSSFRTTQSGLSLNIDVS 275

Qy        255 THKVMRTETIYDIMRRCSHNPARHQDEV-----RVNVLDLIVLTDYNNRTYRINDVDFGQ 309
              |  ::    : | :   |:  |   ::      :  : :| : |  :|: ::|: :
Db        276 TTMIISPGPVVDFL--ISNQNVRDPFQLDWAKAKRTLKNLRIKTSPSNQEFKISGLSELP 333

Qy        310 TPKSTFSCKGR----------DISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASE 359
              : ||: ||:           :|:  :|::    | :|       |:  : |
Db        334 CREQTFTLKGKGGGDGEDGNEEITVYDYFVKVRKIDLRYSADLPCINVGKPKR------- 386

Qy        360 LVVLIP-ELCRVTGLNAEMRSNFQLMRA-MSSYTRMNPKQR----TDRLRAFNHRLQNTP 413
                | |:| :  |    ::   | || :   :|  |::|    :| ||  |:  :
Db        387 -PTFFPIEVCELVSLQRYTKALSTLQRASLVEKSRQKPQERMKILSDALRTSNYGAE--- 442

Qy        414 ESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSD 473
                :||:  : :    |||:||::    : | ||:     |  |       |:
Db        443 ---PMLRNCGISISTGFTEVEGRVLPAPRLKFGNGEDLNPRNGRW----NVSRVKFVEPS 495

Qy        474 GLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDD-----RTGTYVRA- 527
              ::|||| |      ::| |:  | |     |: |  |  | ::|:     |    ||
Db        496 KIERWAV-ANFSARCDVRGLVRDLIRIGDMKGITIEQP--FDVFDENPQFRRAPPMVRVE 552

Qy        528 -MDDCVRSD----PKLILCLVPN-DNAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMS 581
               | : ::|     |: :|||:|:  | : |   ||:   |  :  | :       :: |
Db        553 KMFEHIQSKLPGAPQFLLCLLPDRKNCDIYGPWKKKNLADFGIINQCMCPLRVNDQYL-- 610

Qy        582 IATKIAIQLNCKLG---------YTPWMIELPLSGLMTIGFDIAKSTRDRK--RAYGALI 630
               | : :::| |||          ::| :  :  :  : :| |::  :  :     : |::
Db        611 --TNVMLKINAKLGGLNSLLGVEHSPSLPVVSKAPTLILGMDVSHGSPGQTDIPSIAAVV 668

Qy        631 ASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAK----------ALRQYQHEHRKLPSRI 680
              :|      | | : |   ||   : : |: ::::           |  |     |: |  |
Db        669 SSRHWPLISKYRACVRTQSAKMEMIDNLFKLVSEKEDEGIIRELLLDFYTTSGRRKPENI 728

Qy        681 VFYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLSPPQLAYIVVTRSMNTRFFLNG-- 738
              : :|||||     |:   |:  |||  |      :   |:   ||  :: :||||  |
Db        729 IIFRDGVSESQFNQVLNIELDRIIEACKFLDENWE---PKFVVIVAQKNHHTRFFQPGSP 785
```

```
Qy        739 QNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMC 798
              | ||||::|: |  |  |||||  :     ||  || |:||   :| ||:::|:| : :
Db        786 DNVPPGTVIDNKICHPRNYDFYLCAHAGMIGTSRPTHYHVLLDQVGFSPDQLQELVHSLS 845

Qy        799 HLYYNWSGTTRVPAVCQYAKKLATLVG 825
              ::|   :    | |   ||   || :|
Db        846 YVYQRSTTAISVVAPICYAHLAATQLG 872


RESULT 15
US-10-174-363-54
; Sequence 54, Application US/10174363
; Publication No. US20030077623A1
; GENERAL INFORMATION:
;  APPLICANT: Butler, Karlene H.
;  APPLICANT:  Cahoon, Rebecca E
;  APPLICANT:  Harvell, Leslie T.
;  APPLICANT:  Rafalski, Antoni J.
;  APPLICANT:  Sakai, Hajime
;  TITLE OF INVENTION: Polynucleotides And Polypeptides Involved In Post-Transcriptional
;  TITLE OF INVENTION:  Gene Silencing
;  FILE REFERENCE: BB1454 US NA
;  CURRENT APPLICATION NUMBER: US/10/174,363
;  CURRENT FILING DATE:  2002-06-17
;  PRIOR APPLICATION NUMBER: 60/298,973
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 56
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 54
;   LENGTH: 1101
;   TYPE: PRT
;   ORGANISM: Oryza sativa
US-10-174-363-54

  Query Match             10.6%;  Score 465.5;  DB 15;  Length 1101;
  Best Local Similarity   23.2%;  Pred. No. 1e-35;
  Matches  226;  Conservative  154;  Mismatches  380;  Indels  213;  Gaps   45;

Qy          3 DDQGRGRRRPXNEDDSSTSRGS-----GDGPRVKVFRGSSSGDPRADPRIEASR--ERRA 55
              | ||||. ||         ||         |||    |   ||  |  | :  :   : :|
Db        121 DYQGRGGPRP---------RGGMPQPYYGGPRGSGGRSVPSGSSRTVPELHQAPHVQYQA 171

Qy         56 LEEAPRREGGPPERKPWGD--------QYDYLNTRPVELVSK------------------ 89
                :|   |      :|  :         |:  | ||      |:
Db        172 PMVSPTPSGAGSSSQPAAEVSSGQVQQQFQQLATRDQSSTSQAIQIAPPSSKSVRFPLRP 231

Qy         90 -KGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMGV------LSNHANLL 142
               ||| |   ::: | |  :  |:  :  | |   | : :  |   |       |  :::|
Db        232 GKGTYGDRCIVKANHFFAEL-PDKDLHQYDVSITPEVTSRGVNRAVMFELVTLYRYSHLG 290

Qy        143 GSGYLFDGLQL--------FTTRKFEQEITVL---------SGKSKLDIEYKISIKF--- 182
              |    :|| :          | :| |  |||:           |  : :  ::: |||
Db        291 GRLPAYDGRKSLYTAGPLPFASRTF--EITLQDEEDSLGGGQGTQRRERLFRVVIKFAAR 348

Qy        183 ----------VGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPR--AKIEIREFK 230
                       |  : |    ||||:::| | :       |||: : |    : :: |
Db        349 ADLHHLAMFLAGRQADAPQEALQVLDIVL-RELPTTRYSPVGRSFYSPNLGRRQQLGE-G 406

Qy        231 MELWPGYETSIRQHEKDILLGTEITHKV-MRTETIYDIM-----RRCSHNPARHQDEVRV 284
              :| | |:  |||  :  : |  :::    :    : | :     |  |  |    | |::
Db        407 LESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIDFVAQLLNRDISVRPLSDSDRVKI 466

Qy        285 NVL---DLIVLTDYNN--RTYRINDVDFGQTPKSTFSCKGRDISF-----------VEYY 328
                  : :|   |  | |||:    | | ::|     |::||           |:|:
Db        467 KKALRGVKVEVTHRGNMRRKYRIS----GLTSQAT-----RELSFPVDDRGTVKTVVQYF 517

Qy        329 LTKYNIRIRDHNQPLLISKNRDKALKTNASELVVLIPELCR-VTGLNAEMRSNFQLMRAM 387
              |  |   |:    | |   |: :           |  |:|: | |     | | : : |:
Db        518 LETYGFSIQHTTLPCLQVGNQQRP--------NYLPMEVCKIVEGQRYSKRLNEKQITAL 569

Qy        388 SSYTRMNPKQR-TDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFH 446
              |   |::|  | ||  :|   |     :  ::: :::|: :  |: |::    : :|
Db        570 LKVTCQRPQERELDILRTVSH---NAYHEDQYAQEFGIKIDERLASVEARVLPPPRLKYH 626

Qy        447 NGKVPAGENAD-----WQRHFRDQRMLTTPSDGLDRWAVIAPQRN---------SHELRT 492
              :    :|   |       | :  :::|:      :: || |   ||          |||
Db        627 D----SGREKDVLPRVGQWNMMNKKMVN--GGRVNNWACINFSRNVQDSAARGFCHELAI 680

Qy        493 LLDSLYRAASGMGLRIR--------SPQEFIIYDDRTGTYVRAMDDCVRSDPK---LILC 541
              :        |||   :          |:   :        |  || : :|   :   |::
Db        681 MCQ-----ISGMDFALEPVLPPLTARPEH--VERALKARYQDAM-NMLRPQGRELDLLIV 732

Qy        542 LVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLGYTPWMI 601
              ::|::|   |  :|:    |  : :|    |      |   :|  :|:::| |:|    ::
Db        733 ILPDNNGSLYGDLKRICETDLGLVSQCCLTKHVFKMSKQYLA-NVALKINVKVGGRNTVL 791

Qy        602 ------ELPL-SGLMTI--GFDIA--KSTRDRKRAYGALIASMDLQQNSTYFSTVTECSA 650
                    :|| |   ||  | |:       |   :  |::|| |  : : |   |:  :
Db        792 VDALTRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDWPEVTKYAGLVSAQAH 851

Qy        651 FDVLANTL---WP----------MIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFE 697
                  |   |   |          || : | ::    : | ||:|||||| |   |:
Db        852 RQELIQDLFKVWQDPHRGTVTGGMIKELLISFKRATGQKPQRIIFYRDGVSEGQFYQVLL 911

Qy        698 FEVKDIIEKLKTEYARVQLSP---PQLAYIVVTRSMNTRFFLNGQNPP----------PG 744
              :|: | | |     |  | |   | : ::|| :  :|| | |  |            ||
Db        912 YEL-DAIRK-----ACASLEPNYQPPVTFVVVQKRHHTRLFANNHNDQRTVDRSGNILPG 965

Qy        745 TIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNW 804
              |:||  |  | .:|||| |    |||  |  |:||:     : :::| ||  :|: |
Db        966 TVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDENKFTADELQTLTNNLCYTYARC 1025

Qy        805 SGTTRVPAVCQYA 817
              : :  :      ||
Db       1026 TRSVSIVPPAYYA 1038


Search completed: July 22, 2003, 07:07:49
Job time : 23.6828 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 06:25:13 ; Search time 16.7549 Seconds
(without alignments)
4836.881 Million cell updates/sec

Title: US-09-873-737A-2
Perfect score: 4389
Sequence: 1 MADDQGRGRRRPXNEDDSST.....VGTNLHSIPQNALEKKFYYL 843

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1087.5	24.8	824	2	T20351	hypothetical prote
2	953.5	21.7	722	2	T30995	hypothetical prote
3	474	10.8	1000	2	T22391	hypothetical prote
4	465.5	10.6	891	2	T32079	hypothetical prote
5	454.5	10.4	813	2	JC6569	translation initia
6	449	10.2	930	2	A84668	Argonaute (AGO1)-l
7	447	10.2	834	2	T41568	argonaute-like pro
8	433.5	9.9	1032	2	T23164	hypothetical prote
9	433	9.9	997	2	A84678	Argonaute (AGO1)-l
10	433	9.9	1035	2	T23165	hypothetical prote
11	431.5	9.8	988	2	T52134	Zwille protein [im
12	429.5	9.8	1040	2	D88568	protein ZK757.3 [i
13	420.5	9.6	958	2	S41013	hypothetical prote
14	360.5	8.2	1014	2	H86438	protein T19E23.7 [
15	353.5	8.1	887	2	T01113	translation initia
16	351.5	8.0	898	2	T25164	hypothetical prote
17	307	7.0	1139	2	T33275	hypothetical prote
18	303	6.9	945	2	T23965	hypothetical prote
19	295.5	6.7	990	2	F88925	protein T22H9.3 [i
20	289.5	6.6	892	2	T15195	hypothetical prote
21	274	6.2	939	2	T18974	hypothetical prote
22	267	6.1	965	2	T22933	hypothetical prote
23	256.5	5.8	1032	2	T19324	hypothetical prote
24	248	5.7	674	2	T19268	hypothetical prote
25	210	4.8	324	2	G87774	protein C24A11.3 [
26	209.5	4.8	1030	2	T16114	hypothetical prote
27	209	4.8	944	2	T18911	hypothetical prote
28	207	4.7	1057	2	T16676	hypothetical prote
29	205.5	4.7	905	2	T23510	hypothetical prote
30	202	4.6	363	2	T34486	hypothetical prote
31	199	4.5	567	2	T34339	hypothetical prote
32	190	4.3	706	2	H70425	conserved hypothet
33	166.5	3.8	881	2	T31818	hypothetical prote
34	160.5	3.7	1819	2	D97033	uncharacterized pr
35	150.5	3.4	713	2	H64464	hypothetical prote
36	148.5	3.4	697	2	T15179	hypothetical prote
37	132.5	3.0	185	2	T27784	hypothetical prote
38	130.5	3.0	1228	2	T18897	hypothetical prote
39	125.5	2.9	2028	2	T08025	DNA-directed RNA p
40	122.5	2.8	1191	2	B97116	chromosome segrega
41	122.5	2.8	5147	1	IJFFTM	cadherin-related t
42	121.5	2.8	1092	2	T18354	adhesin - Mycoplas
43	121.5	2.8	1312	2	T30845	probable DNA repai
44	121.5	2.8	2022	2	T43214	ovtl protein - nem
45	119	2.7	816	2	B86285	hypothetical prote

ALIGNMENTS

RESULT 1
T20351
hypothetical protein D2030.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T20351
R;Wilkinson, J.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19261
A;Accession: T20351
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-824 <WIL>
A;Cross-references: EMBL:Z73906; PIDN:CAA98113.1; GSPDB:GN00019; CESP:D2030.6
A;Experimental source: clone D2030
C;Genetics:
A;Gene: CESP:D2030.6
A;Map position: 1
A;Introns: 118/3; 223/3; 304/3; 452/1; 573/3; 632/1; 711/2; 773/3

Query Match 24.8%; Score 1087.5; DB 2; Length 824;
Best Local Similarity 29.9%; Pred. No. 1.6e-67;
Matches 267; Conservative 167; Mismatches 341; Indels 117; Gaps 22;

```
Qy          1 MADDQGRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAP 60
              ||   ||||            ||||          ::||
Db          1 MASGSGRGR----------GRGSG---------SNNSG--------------------- 19

Qy         61 RREGGPPERKPWGDQYDYLNTRPVELV------SKKGTDGVPVMLQTNFFRXK-TKPEWR 113
                        |    || |   :|       || |: | |     ||   : |: ::
Db         20 ------------GKDQKYLGTIQPDLFIRQQGQSKTGSSGQPQKCFANFIPIEMTQSDYS 67

Qy        114 IVHYHVEFVPTIENPRVRMGVL--SNHANLLGSGYLFDGLQLFTTRKFEQ----EITVLS 167
              |  ||||| ||:::   |  :|  :|  : :|  ::|||: |:   ::||    |:
Db         68 IYQYHVEFEPTVDSKANREKMLRDNNVTDEIGHHFVFDGMILYLKEEWEQNQMIEVQHPI 127

Qy        168 GKSKLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLF---DPRAKI 224
               :| : | :| : :|:      :|: : : | |:|||   | |  :||| |   | ||
Db        128 DRSLICIRFKQTNRFL----VDDPQTINIFNTIIRRSFDALQLTQLGRNYFNWGDSRA-- 181

Qy        225 EIREFKMELWPGYETSIRQHEKDILLGTEITHKVMRTETIYDI----MRRCSHNPARHQD 280
               : :: | : |||||:|| :|:: :|  |   |::| |::| :    :| | :|| | |:
Db        182 -VPDYNMSILPGYETAIRMYEENFMLCVENRFKMVREESMYILFHKELRSCQNNPQRVQE 240

Qy        281 EVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHN 340
              ::       ::| |||: :|   :|:  :| | |   |: |:  ||:  :| | |   :
Db        241 KMNEMYGGTTIITRYNNKLHRYTRLDYSISPLSEFVKDGQSITLKEYFKNQYGIEITVDD 300

Qy        341 QPLLISKNRDKALKTNASELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTD 400
              ||::||: : |       ::  ::|||| ||| ||| :|::|: :: :|||:|:||
Db        301 QPIIISEGKPKQ-PGEPPQVSYIVPELCFPTGLTDEMRKDFKMMKEIAKHTRMSPQQRLV 359
```

```
Qy        401 RLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQR 460
Db        360 ESRKLIVDLSKNEKVMECFKYWGISLGQDLANVQARVLKSEPL--QGKKTYEGKQAEWAR 417

Qy        461 HFRDQRMLTTPSDGLDRWAVIAPQRNSHEL--RTLLDSLYRAASGMGLRIRSPQEFIIYD 518
Db        418 GVKECGIYR--GSNMTNWIVIGPGSGNSGLLSQKFIEEARRLGKILQVQLGEPMCVPIRG 475

Qy        519 DRTGTYVRAMDDCVR----SDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT 574
Db        476 ISPNDYLEGVKGAIKQVDGEDIHMLVVMLADDNKTRYDSLKKFLCVECPIPNQCVNLRTL 535

Qy        575 --------KNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAY 626
Db        536 AGKSKDGGENKNLGSIVLKIVLQMICKTGGALWKVNIPLKNTMIVGYDLYHDSTLKGKTV 595

Qy        627 GALIA--SMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEH-RKLPSRIVFY 683
Db        596 GACVSTTSNDFTQ---FYSQTRPHENPTQLGNNLTHFVRKALKQYYDSNDQTLPSRLILY 652

Qy        684 RDGVSSGSLKQLFEFEVK------DIIEKLKTEYARVQLSPPQLAYIVVTRSMNTRFFLN 737
Db        653 RDGAGDGQIPYIKNTEVKLVRDACDAVTDKAAELSNKVQEKIKLAFIIVTKRVNMRILKQ 712

Qy        738 GQ------NPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQ 791
Db        713 GSSLDNAINPQPGTVVDTTVTRPERMDFYLVPQFVNQGTVTPVSYNIIHDDTDLGPDKHQ 772

Qy        792 KLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKFYYL 843
Db        773 QLAFKLCHLYYNWQGTVRVPAPCQYAHKLAFLTAQSLHDDANGCLRDKLFFL 824


RESULT 2
T30995
hypothetical protein C01G5.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 29-Oct-1999
C;Accession: T30995
R;Bradshaw, H.; Stellyes, L.
submitted to the EMBL Data Library, August 1999
A;Description: The sequence of C. elegans cosmid C01G5.
A;Reference number: Z20956
A;Accession: T30995
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-722 <BRA>
A;Cross-references: EMBL:U50068; PIDN:AAB37734.1
A;Experimental source: strain Bristol N2
C;Genetics:
A;Map position: IV
A;Introns: 58/3; 121/3; 202/3; 350/1; 471/3; 530/1; 609/2; 671/3
A;Note: C01G5.2

  Query Match             21.7%;  Score 953.5;  DB 2;  Length 722;
  Best Local Similarity   29.8%;  Pred. No. 2.9e-58;
  Matches  230;  Conservative  144;  Mismatches  306;  Indels   91;  Gaps   19;

Qy        111 EWRIVHYHVEFVPTIENPRVRMGVLSNHANL--LGSGYLFDGLQLFTTRKFEQ----EIT 164
Db          5 DYSIYQYHVEFEPTVDSKATRENMLRQPSVTVEIGKHFVFDGMILYLKEEWDQNQMIEVQ 64

Qy        165 VLSGKSKLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKI 224
Db         65 HPNDNSLICIRFKKTNRFL----VDDPQTINIFNTIIRRSFDAMKLTQIGRNYFD----- 115

Qy        225 EIREFKMELWPGYETSIRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARHQDEVRV 284
Db        116 ---------WD------------------NSRALRKE-----LRSCQNNRQRVQEKMNE 142

Qy        285 NVLDLIVLTDYNNRTYRINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLL 344
Db        143 VYGGSTIITRYNNKLHRFTRLDNEITPLSKFQKDGEQIILKEYFKNQYDIDITDDEQFII 202

Qy        345 ISKNRDKALKTNASELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRA 404
Db        203 ISEGKPKQ-PGEPPQVNYIVPELCFPTGLTDEMRKDFKMMKEIAKHTRMSPQQRLDETRK 261

Qy        405 FNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVP-AGENADWQRHFR 463
Db        262 LITKLSQNQTMMECFQYWGISLGQDLANVQARVLKSEPL---QGKRQYEGKQAEWARGVK 318

Qy        464 DQRMLTTPSDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMG----LRIRSPQEFIIYDD 519
Db        319 ECGIYR--GSNMTNWIVIGPGSGNSGL--LAQKFIAEARNLGRTLQVQLGEPMCVKINGI 374

Qy        520 RTGTYVRAMDDCVRS----DPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT- 574
Db        375 SPNDYLEGLKAAIKSVDGEEIHMLVVMLADDNKTRYDSLKKYLCVECPIPNQCVNLRTLA 434

Qy        575 -------KNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYG 627
Db        435 GKSKDGGENKNLGSIVLKIVLQMICKTGGALWKVNIPLKSTMIVGYDLYHDSTLKGKTVG 494

Qy        628 ALIA--SMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRK-LPSRIVFYR 684
Db        495 ACVSTTSNDFTQ---FYSQTRPHENPTQLGNNLTHFVRKSLKQYYDNNDKTLPSRLILYR 551

Qy        685 DGVSSGSLKQLFEFEVK------DIIEKLKTEYARVQLSPPQLAYIVVTRSMNTRFFLNG 738
Db        552 DGAGDGQIPYIKNTEVKLVRDACDAVTDKAAELSNKVQEKIKLAFIIVTKRVNMRILKQG 611

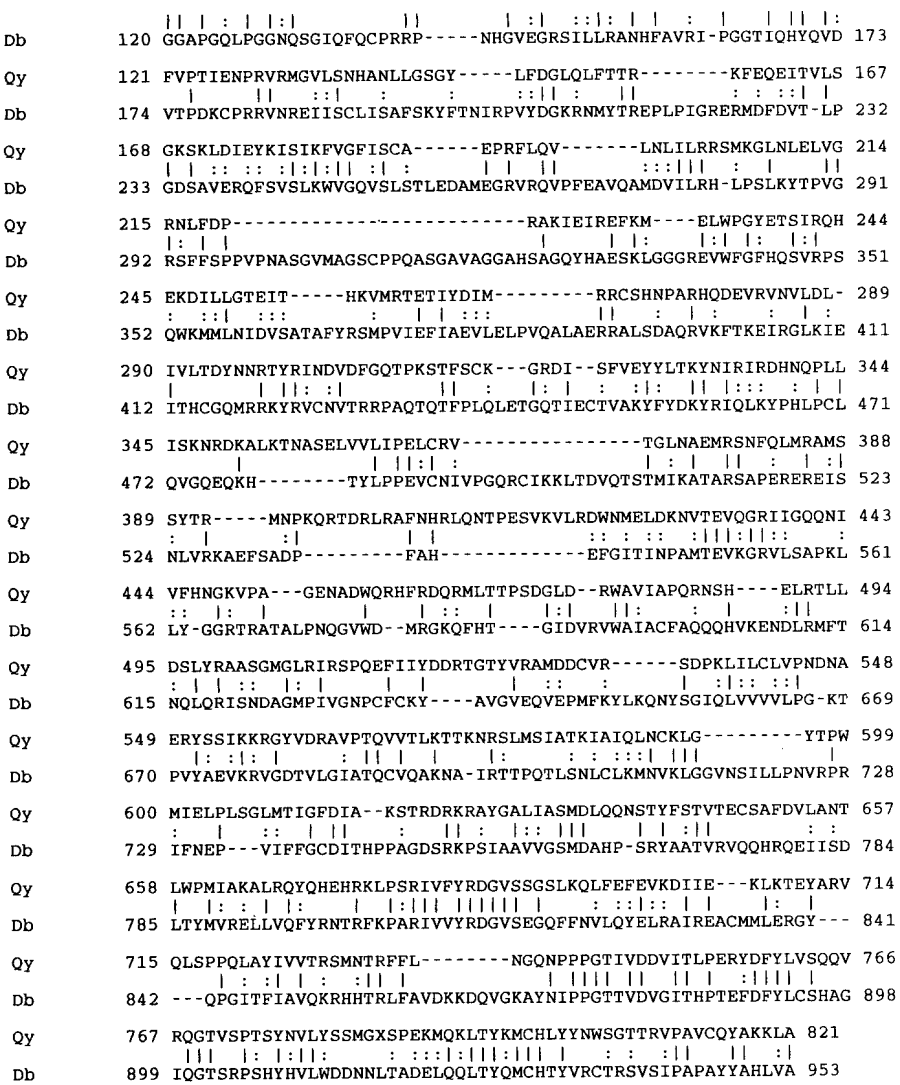
Qy        739 Q------NPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQK 792
Db        612 SSSKSAINPQPGTVVDTTVTRPERMDFYLVPQFVNQGTVTPVSYNIIHDDTGLGPDKHQQ 671

Qy        793 LTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKFYYL 843
Db        672 LAFKLCHLYYNWQGTVRVPAPCQYAHKLAFLTAQSLHDDANGYLRDKLFFL 722


RESULT 3
T22391
hypothetical protein F48F7.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T22391; T23234
R;Coles, L.
submitted to the EMBL Data Library, February 1996
A;Reference number: Z19559
A;Accession: T22391
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1000 <WIL>
A;Cross-references: EMBL:Z69661; PIDN:CAA93496.1; GSPDB:GN00028; CESP:F48F7.1
A;Experimental source: clone F48F7
R;Coles, L.
submitted to the EMBL Data Library, February 1996
A;Reference number: Z19712
A;Accession: T23234
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1000 <WI2>
A;Cross-references: EMBL:Z69663; PIDN:CAA93512.1; GSPDB:GN00028; CESP:F48F7.1
A;Experimental source: clone K02B9
C;Genetics:
A;Gene: CESP:F48F7.1
A;Map position: X
A;Introns: 70/3; 128/2; 185/2; 673/1
C;Superfamily: rabbit translation initiation factor eIF-2C

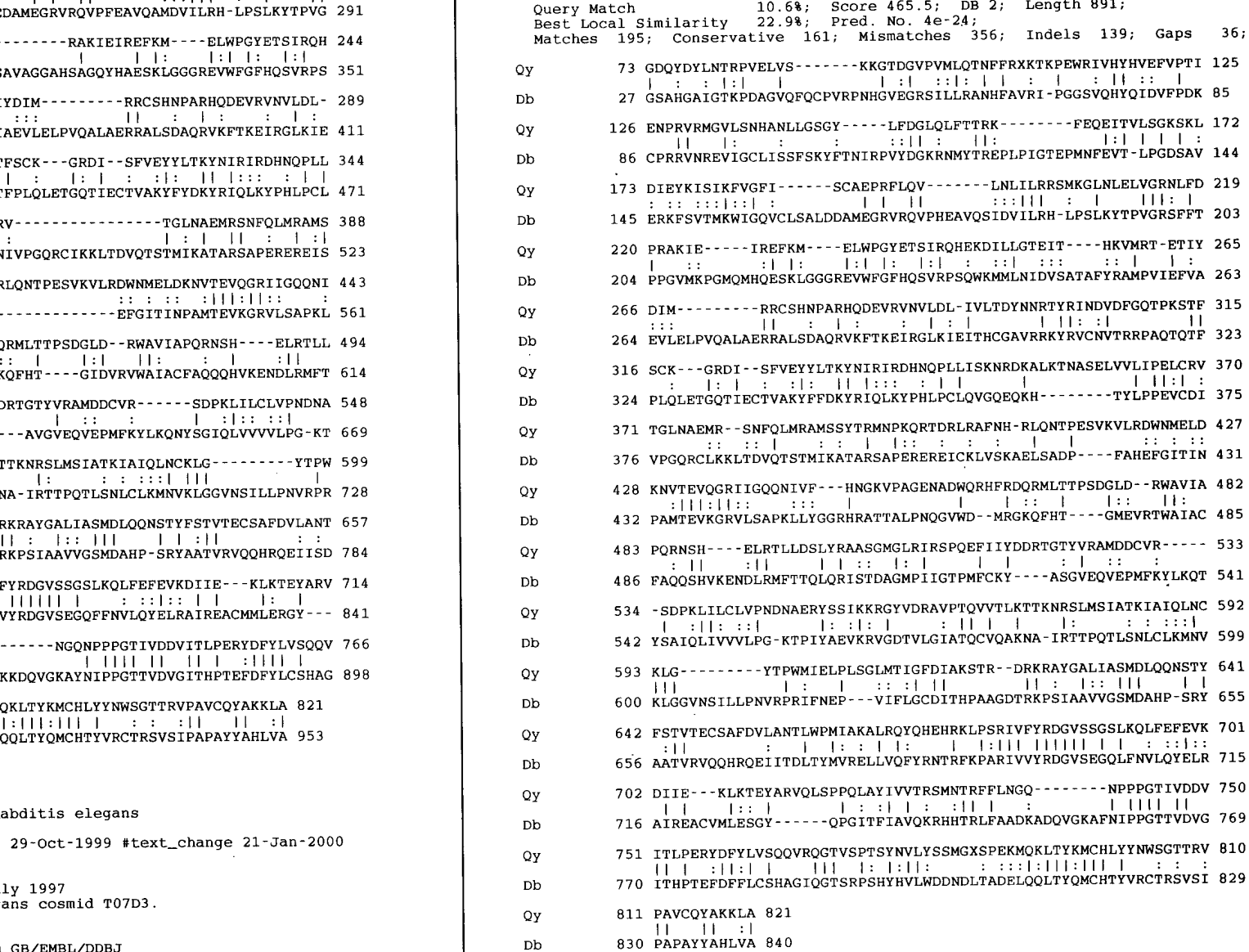
  Query Match             10.8%;  Score 474;  DB 2;  Length 1000;
  Best Local Similarity   23.4%;  Pred. No. 1.2e-24;
  Matches  209;  Conservative  148;  Mismatches  340;  Indels  198;  Gaps   38;

Qy         64 GGPPERKPWGDQYD---YLNTRPVELVSKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVE 120
```



RESULT 4
T32079
hypothetical protein T07D3.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C;Accession: T32079
R;Lamar, B.; Wamsley, P.; Twyman, B.
submitted to the EMBL Data Library, July 1997
A;Description: The sequence of C. elegans cosmid T07D3.
A;Reference number: Z21121
A;Accession: T32079
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-891 <LAM>
A;Cross-references: EMBL:AF016682; PIDN:AAB66187.1; GSPDB:GN00020; CESP:T07D3.7
A;Experimental source: strain Bristol N2; clone T07D3
C;Genetics:
A;Gene: CESP:T07D3.7
A;Map position: 2
A;Introns: 34/2; 92/2; 206/1; 560/1; 879/3
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 10.6%; Score 465.5; DB 2; Length 891;
Best Local Similarity 22.9%; Pred. No. 4e-24;
Matches 195; Conservative 161; Mismatches 356; Indels 139; Gaps 36;



RESULT 5
JC6569
translation initiation factor eIF-2C - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 21-Jul-2000
C;Accession: PC6505; JC6569
R;Zou, C.; Zhang, Z.; Wu, S.; Osterman, J.C.
Gene 211, 187-194, 1998
A;Title: Molecular cloning and characterization of a rabbit eIF2C protein.
A;Reference number: JC6569; MUID:98267198; PMID:9602122
A;Accession: PC6505
A;Molecule type: protein
A;Residues: 336-359;380-409;694-711 <ZOU1>
A;Experimental source: liver
A;Accession: JC6569
A;Molecule type: mRNA
A;Residues: 1-813 <ZOU2>
A;Cross-references: GB:AF005355; NID:g3253158; PIDN:AAC24323.1; PID:g3253159
C;Superfamily: rabbit translation initiation factor eIF-2C
C;Keywords: liver

Query Match 10.4%; Score 454.5; DB 2; Length 813;
Best Local Similarity 22.7%; Pred. No. 2e-23;
Matches 193; Conservative 142; Mismatches 289; Indels 225; Gaps 38;

```
Qy        110 PEWRIVHYHVEFVPTIENPRVRMGVLSN-----HANLLGSGY-LFDGLQ-LFTT------ 156
Db          4 PKIDIYHYELDIKPEKCPRRVNREIVEHMVQHFKAQIFGDRKPVFDGRKNLYTAMPLPIG 63

Qy        157 -RKFEQEITVLSGKSKLDIEYKISIKFVGFISC------------AEP-RFLQVLNLILR 202
Db         64 REKVELEVT-LPGEGK-DRIFKVSIKWVSCVSLQALHDALSGRLPSVPFETIQALDVVMR 121

Qy        203 RSMKGLNLELVGRNLF-------DPRAKIEIREFKMELWPGYETSIR------------- 242
Db        122 H-LPSMRYTPVGRSFFTASEGCSNPLGG------GREVWFGFHQSVRPSLWKMMLNIDVS 174

Qy        243 ------------------------QHEKDIL------------LGTEITH--KVMRTET 263
Db        175 ATAFYKAQPVIEFVCEVLDFKSIEEQQKPLTDSQRVKFTKEIKGLKVEITHCGQMKRKYR 234

Qy        264 IYDIMRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTFSCKGRDIS 323
Db        235 VCNVTRR----PASHQ-------------------TFPLQQ-ESGQTVECTVA------- 263

Qy        324 FVEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASELVVLIP-ELCRVTG-------LNA 375
Db        264 --QYFKDRHKLVLRYPHLP---------CLQVGQEQKHTYLPLEVCNIVAGQRCIKKLTD 312

Qy        376 EMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQG 435
Db        313 NQTSTMIRATARSAPDRQEEISKLMRSASFN--------TDPYVREFGIMVKDEMTDVTG 364

Qy        436 RIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDG--------------LDRWAV- 480
Db        365 RVLQPPSILY------GGRN----------KAIATPVQGVWDMRNKQFHTGIEIKVWAIA 408

Qy        481 -IAPQRNSHE--LRTLLDSLYRAASGMGLRIRSPQEFIIY---DDRTGTYVRAMDDCVRS 534
Db        409 CFAPQRQCTEVHLKSFTEQLRKISRDAGMPIQGQPCFCKYAQGADSVGPMFRHLKN-TYA 467

Qy        535 DPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKL 594
Db        468 GLQLVVVILPG-KTPVYAEVKRVGDTVLGMATQCVQMKNVQ-RTTPQTLSNLCLKINVKL 525

Qy        595 GYT---------PWMIELPLSGLMTIGFDIA--KSTRDRKRAYGALIASMDLQQNSTYFS 643
Db        526 GGVNNILLPQGRPPVFQQP---VIFLGADVTHPPAGDGKKPSIAAVVGSMDAHPN-RYCA 581

Qy        644 TVTECSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDI 703
Db        582 TVRVQQHRQEIIQDLAAMVRELLIQFYKSTRFKPTRIIFYRDGVSEGQFQQVLHHELLAI 641

Qy        704 IE---KLKTEYARVQLSPPQLAYIVVTRSMNTRFFLNGQNP--------PPGTIVDDVIT 752
Db        642 REACIKLEKDY------QPGITFIVVQKRHHTRLFCTDKNERVGKSGNIPAGTTVDTKIT 695

Qy        753 LPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPA 812
Db        696 HPTEFDFYLCSHAGIQGTSRPSHYHVLWDDNRFSSDELQILTYQLCHTYVRCTRSVSIPA 755

Qy        813 VCQYAKKLA 821
Db        756 PAYYAHLVA 764
```

RESULT 6
A84668
Argonaute (AGO1)-like protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: A84668
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: A84668
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-930 <STO>
A;Cross-references: GB:AE002093; NID:g3885334; PIDN:AAC77862.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g27040
A;Map position: 2
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 10.2%; Score 449; DB 2; Length 930;
Best Local Similarity 21.2%; Pred. No. 6e-23;
Matches 204; Conservative 172; Mismatches 351; Indels 233; Gaps 40;

```
Qy         24 SGDGPRVKVFRGSS-SG--------DPRADPRIEASRERRALEEAPRREGGPPERKPWGD 74
Db          5 NGNGADLESANGANGSGVTEALPPPPPVIPPNVEPVRVKTELAE---KKG--PVRVP--- 56

Qy         75 QYDYLNTRPVELVSKK--GTDGVPVMLQTNFFRXK-TKPEWRIVHYHVEFVPTIENPRVR 131
Db         57 ------------MARKGFGTRGQKIPLLTNHFKVDVANLQGHFFHYSVALFYDDGRPVEQ 104

Qy        132 MGV---------LSNHANLLGSGYLFDGLQ-LFT-----TRKFE-----QEITVLS---- 167
Db        105 KGVGRKILDKVHQTYHSDLDGKEFAYDGEKTLFTYGALPSNKMDFSVVLEEVSATSKDFV 164

Qy        168 -------------------------GKSK---LDIEY--KISIKFV-----GFISCAEP 191
Db        165 SRANGNGSPNGNESPSDGDRKRLRRPNRSKNFRVEISYAAKIPLQALANAMRGQESENSQ 224

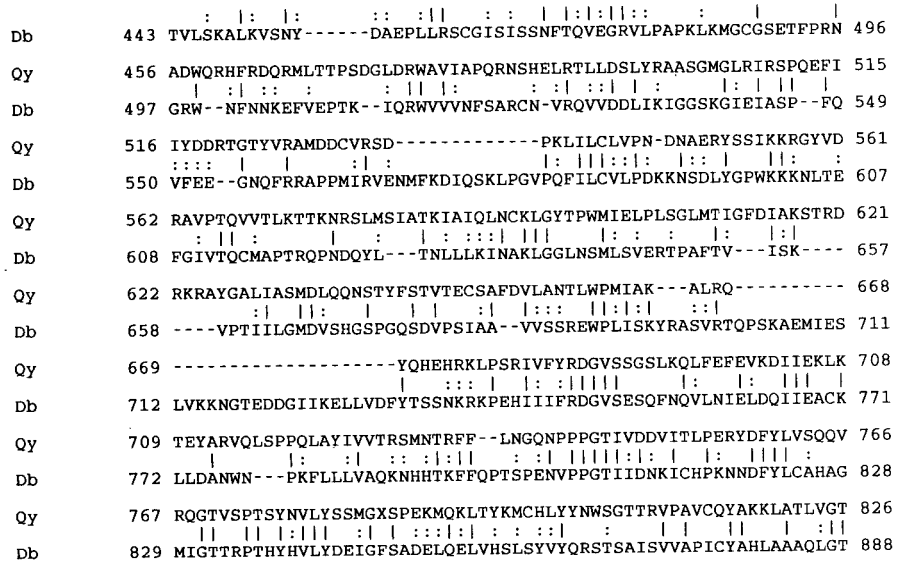
Qy        192 RFLQVLNLILRR--SMKG-------------LNLELVGRNLFDPRAKIEIREFKMELWPG 236
Db        225 EAIRVLDIILRQHAARQGCLLVRQSFFHNDPTNCEPVGGNILGCR-------------G 270

Qy        237 YETSIRQHEKDILLGTEI-THKVMRTETIYDIMRRCSHNPARHQ-----DEVRVNVLDLI 290
Db        271 FHSSFRTTQGGMSLNMDVTTTMIIKPGPVVDFL--IANQNARDPYSIDWSKAKRTLKNLR 328

Qy        291 VLTDYNNRTYRINDVDFGQTPKSTFSCKGR----------DISFVEYYLTKYNIRIRDH 339
Db        329 VKVSPSGQEFKITGLSDKPCREQTFELKKRNPNENGEFETTEVTVADYFRDTRHIDLQYS 388

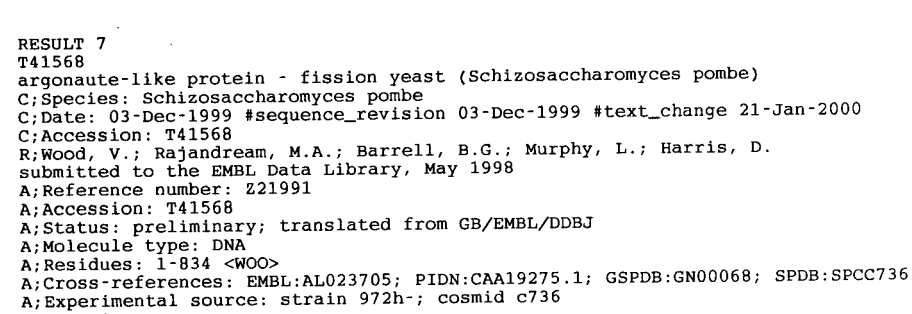
Qy        340 NQPLLISKNRDKALKTNASELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR- 398
Db        389 ADLPCINVGKPKRPTYIPLELCALVPLQRYTKALTTFQRS------ALVEKSRQKPQERM 442

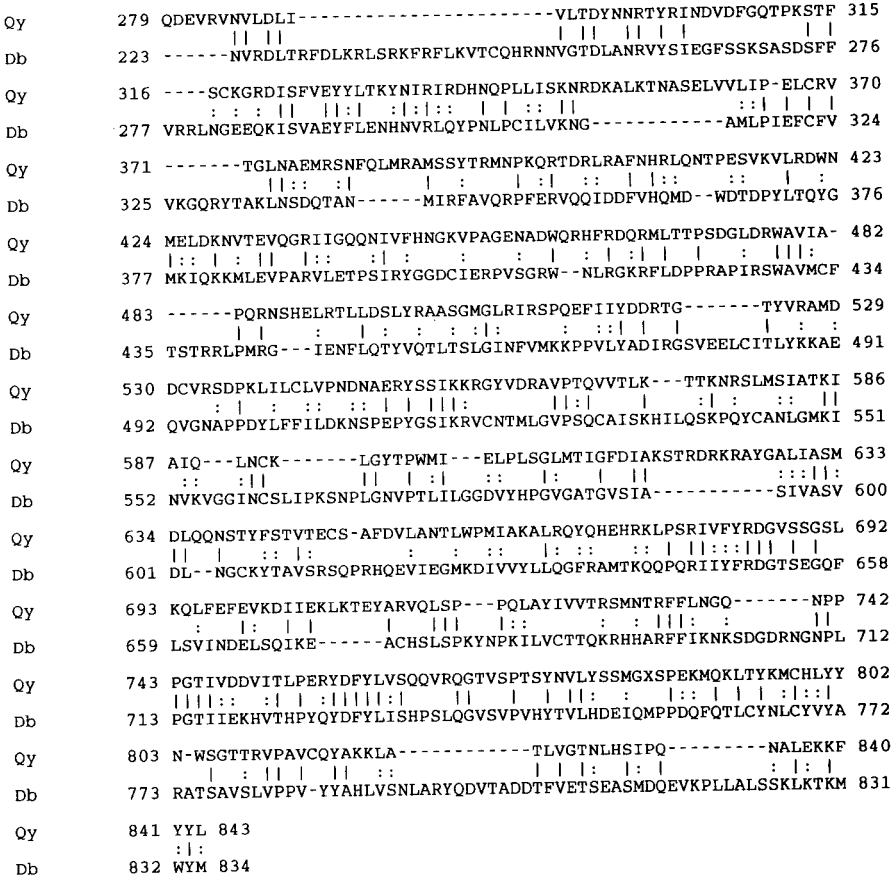
Qy        399 ---TDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN 455
```



RESULT 7
T41568
argonaute-like protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C;Accession: T41568
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.
submitted to the EMBL Data Library, May 1998
A;Reference number: Z21991
A;Accession: T41568
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-834 <WOO>
A;Cross-references: EMBL:AL023705; PIDN:CAA19275.1; GSPDB:GN00068; SPDB:SPCC736.11
A;Experimental source: strain 972h-; cosmid c736
C;Genetics:
A;Gene: SPDB:SPCC736.11
A;Map position: 3
A;Introns: 43/3
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 10.2%; Score 447; DB 2; Length 834;
Best Local Similarity 21.5%; Pred. No. 6.9e-23;
Matches 194; Conservative 157; Mismatches 332; Indels 220; Gaps 37;





RESULT 8
T23164
hypothetical protein T22B3.2a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T23164; T25099
R;Cottage, A.
submitted to the EMBL Data Library, January 1996
A;Reference number: Z19701
A;Accession: T23164
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1032 <WIL>
A;Cross-references: EMBL:Z68750; PIDN:CAA92969.1; GSPDB:GN00022; CESP:T22B3.2a
A;Experimental source: clone K01A6
R;Lennard, N.
submitted to the EMBL Data Library, December 1995
A;Reference number: Z19981
A;Accession: T25099
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1032 <WI2>
A;Cross-references: EMBL:Z68300; PIDN:CAA92618.1; GSPDB:GN00022; CESP:T22B3.2a
A;Experimental source: clone T22B3

C;Genetics:
A;Gene: CESP:T22B3.2a
A;Map position: 4
A;Introns: 19/1; 71/3; 129/2; 424/2; 496/3; 607/1; 666/1; 753/3; 795/2; 961/2; 1003/1
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 9.9%; Score 433.5; DB 2; Length 1032;
Best Local Similarity 20.6%; Pred. No. 8.5e-22;
Matches 211; Conservative 162; Mismatches 368; Indels 281; Gaps 43;

```
Qy          9 RRRPXNEDDSSTSRGSGDG---PRVKVFRGSSSGDPRA-----DPRIEASRERRALEEAP 60
Db          3 RRNATNFVDNNTLTSSGSGSLSPPITSRPASGQASPLSSNGSLSPPVDDQGSVSYNSDSP 62

Qy         61 RREGGPPERKP-WGDQYDYLNTRPVELVSKKGTDGVPVMLQTNFFRXKTK-PEWRIVHYH 118
Db         63 R------DLSPLLLSELACLNMREVVARPGLGTIGRKIPVKSNFFAVDLKNPKMVVVQYH 116

Qy        119 VEFVPTIENPRVR-----------MGVLSNHANLLGSGY--LFDGL-QLFTTRKFEQEIT 164
Db        117 VE----VHHPGCRKLDKDEMRIIFWKAVSDHPNIFHNKFALAYDGAHQLYTVARL--EFP 170

Qy        165 VLSGKSKLDIE-----------YKISIKFVGFISCAEPRF-----------LQVLNLI 200
Db        171 DDQGSVRLDCEATLPKDNRDRTRCAISIQNVGPVLLEMQRTRTNNLDERVLTPIQILDII 230

Qy        201 LRRSM----------------------KGLNLELVGRNLFDPRAKIEIREFKMELWPG 236
Db        231 CRQSLTCPLLKNSANFYTWKSSCYRIPTAAGQALDLEGGK--------------EMWTG 275

Qy        237 YETS--IRQHEKDILLGTEITH------KVMRTETIYDIMRRCSHNPARH---------- 278
Db        276 FFSSAHIASNYRP-LLNIDVAHTAFYKTRITVLQFMCDVLNERTSKPNRNNPRGPGAPGG 334

Qy        279 ------------------------QDEVRVNVLDLIV--------LTDYNNRTY---- 300
Db        335 YRGGRGGARGGSYQNFGNRGPPGANVRDDFGGNGLTFTMDTLSRDTQLSSFETRIFGDSI 394

Qy        301 -------------RINDVDFGQTPKSTFSCKGRD-------ISFVEYYLTKYNIRIRDH 339
Db        395 RGMKIRATHRPNAIRVYKVNSLQLPADKLMFQGIDEEGRQVVCSVADYFSEKYG------ 448

Qy        340 NQPLLISKNRDKALKTNASELVVLIP-ELCRVTGLNAEMRSNFQLMRAMSS----YTRMN 394
Db        449 --PLKYPKL--PCLHVGPPTRNIFLPMEHCLI---DSPQKYNKKMTEKQTSAIIKAAAVD 501

Qy        395 PKQRTDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGE 454
Db        502 ATQREDRIKQL--AAQASFGTDPFLKEFGVAVSSQMIETSARVIQPPPIMF------GGN 553

Qy        455 NADWQRHFRDQRMLTTPSDG---LDRWAVIAP-------------QRNSHELRTLLDSLY 498
Db        554 N-------RSINPVVFPKDGSWSMDHQTLYMPATCRSYSMIALVDPRDQTSLQTFCQSLT 606

Qy        499 RAASGMGLRIRSPQEFIIY----DDRTGTYVRAMD---------DCVRSDPKLILCLVPN 545
Db        607 MKATAMGMNFPRWPDLVKYGRSKEDVCTLFTEIADEYRVTNTVCDC-------IIVVLQS 659

Qy        546 DNAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLG---------- 595
Db        660 KNSDIYMTVKEQSDIVHGIMSQCVLMKNV-SRPTPATCANIILKLNMKMGGINSRIVADQ 718

Qy        596 -YTPWMIELPLSGLMTIGFDIAKSTRDRKR----AYGALIASMDLQQNSTYFSTVTECSA 650
Db        719 ITNKYLVDQP---TMVVGIDVTHPTQAEMRMNMPSVAAIVANVDLLPQSYGANVKVQKKC 775

Qy        651 FDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTE 710
Db        776 RESVVYLLDAIRERIITFYRHTKQK-PARIIVYRDGVSEGQFSEV----LREEIQSIRTA 830

Qy        711 YARV--QLSPPQLAYIVVTRSMNTRFFL--------NGQNPPPGTIVDDVITLPERYDFY 760
Db        831 CLAIAEDFRPP-ITYIVVQKRHHARIFCKFPNDMVGKAKNVPPGTTVDTGIVSPEGFDFY 889

Qy        761 LVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKL 820
Db        890 LCSHYGVQGTSRPARYHVLLDECKFTADEIQNITYGMCHTYGRCTRSVSIPTPVYYADLV 949

Qy        821 AT 822
Db        950 AT 951
```

RESULT 9
A84678
Argonaute (AGO1)-like protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: A84678
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: A84678
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-997 <STO>
A;Cross-references: GB:AE002093; NID:g4510428; PIDN:AAD21514.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g27880
A;Map position: 2
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 9.9%; Score 433; DB 2; Length 997;
Best Local Similarity 21.7%; Pred. No. 8.7e-22;
Matches 228; Conservative 161; Mismatches 377; Indels 286; Gaps 47;

```
Qy          5 QGRGRRRPXNEDDS-----------------STSRGSGDGPRVKVFRGSSSGDPRA--- 43
Db         16 RGGGRRSDQRQDQSSGQVAWPGLQQSYGGRGGSVSAGRGRG---NVGRGENTGDLTATQV 72

Qy         44 ----------------DPRI--------------EASRERRALEEAPRREGGPPERKP 71
Db         73 PVASAVSGGRGRGNIGDPTFSVASSSKTVSVASSSKEESKNTEVSETMSNLQITSTETKP 132

Qy         72 WGDQYDYLNTRPVELVSK--KGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPR 129
Db        133 EMTSLPPASSKAVTFPVRPGRGTLGKKVMVRANHFLVQV-ADRDLYHYDVSINPEVISKT 191

Qy        130 VRMGVLS------NHANLLGSGYLFDGLQ-LFTT--RKFEQEITVLS-------GKSKLD 173
Db        192 VNRNVMKLLVKNYKDSHLGGKSPAYDGRKSLYTAGPLPFDSKEFVVNLAEKRADGSSGKD 251

Qy        174 IEYKISIKFVGFISCAE-PRFL-----------QVLNLILRRSMKGLNLELVGRNLFDP 220
Db        252 RPFKVAVKNVTSTDLYQLQQFLDRKQREAPYDTIQVLDVVLRDKPSN-DYVSVGRSFFHT 310

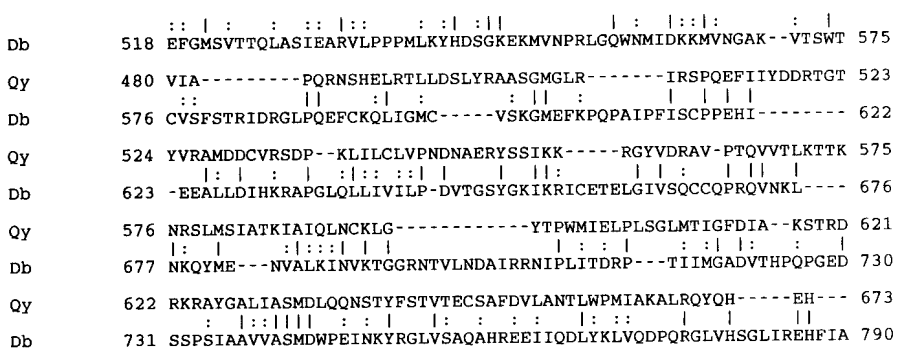
Qy        221 RAKIEIREFKMEL------WPGYETSIRQHEKDILLGTEITHK-----VMRTETIYDIMR 269
Db        311 SLGKDARDGRGELGDGIEYWRGYFQSLRLTQMGLSLNIDVSARSFYEPIVVTDFISKFLN 370

Qy        270 -RCSHNPARHQDEVRV-NVLDLIVLTDYNNRTYRINDVDFGQTPKSTF----SCKGRDIS 323
Db        371 IRDLNRPLRDSDRLKVKKVL----------RTLKVKLLHWNCTKSAKISGISSLPIRELR 420

Qy        324 F----------VEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASELVVLIP-ELCRVTG 372
Db        421 FTLEDKSEKTVVQYFAEKYNYRVKYQALP---------AIQTGSDTRPVYLPMELCQID- 470

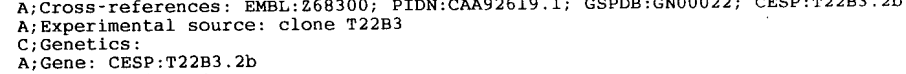
Qy        373 LNAEMRSNFQLMRAMSSYT-RMNPKQRTDRLRAFNHRLQNTPESVKVL----------R 420
Db        471 ------------EGQRYTKRLNEKQVTALLKATCQRPPDRENSIKNLVVKNNYNDDLSK 517

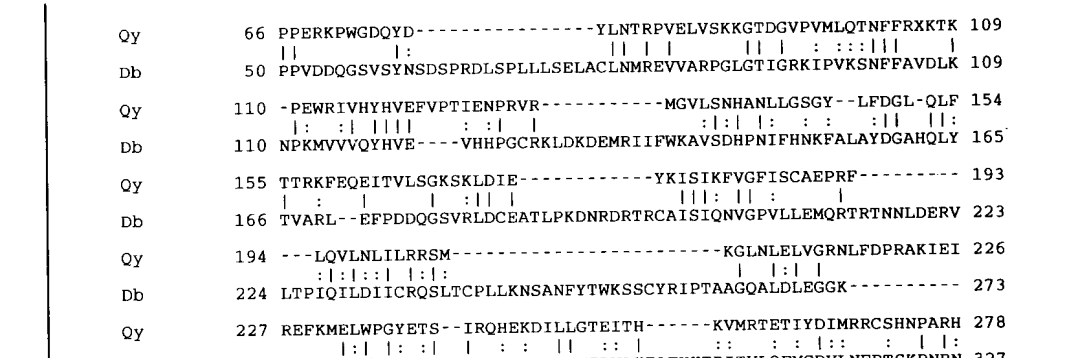
Qy        421 DWNMELDKNVTEVQGRIIGQQNIVFH-NGKVPAGENADWQRHFRDQRMLTTPSDGLDRWA 479
```



RESULT 10
T23165
hypothetical protein T22B3.2b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T23165; T25100
R;Cottage, A.
submitted to the EMBL Data Library, January 1996
A;Reference number: Z19701
A;Accession: T23165
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1035 <WIL>
A;Cross-references: EMBL:Z68750; PIDN:CAA92970.1; GSPDB:GN00022; CESP:T22B3.2b
A;Experimental source: clone K01A6
R;Lennard, N.
submitted to the EMBL Data Library, December 1995
A;Reference number: Z19981
A;Accession: T25100
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1035 <WI2>
A;Cross-references: EMBL:Z68300; PIDN:CAA92619.1; GSPDB:GN00022; CESP:T22B3.2b
A;Experimental source: clone T22B3
C;Genetics:
A;Gene: CESP:T22B3.2b
A;Map position: 4
A;Introns: 19/1; 74/3; 132/2; 427/2; 499/3; 610/1; 669/1; 756/3; 798/2; 964/2; 1006/1
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 9.9%; Score 433; DB 2; Length 1035;
Best Local Similarity 20.6%; Pred. No. 9.2e-22;
Matches 213; Conservative 158; Mismatches 363; Indels 298; Gaps 43;





RESULT 11
T52134
Zwille protein [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Nov-2000
C;Accession: T52134
R;Laux, T.
submitted to the EMBL Data Library, January 1998
A;Reference number: Z25975
A;Accession: T52134
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-988 <LAU>
A;Cross-references: EMBL:AJ223508; PIDN:CAA11429.1
C;Genetics:
A;Gene: zwille
A;Map position: 5
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 9.8%; Score 431.5; DB 2; Length 988;
Best Local Similarity 23.3%; Pred. No. 1.1e-21;
Matches 197; Conservative 139; Mismatches 346; Indels 163; Gaps 37;

```
Qy         91 GTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMGVLSNHANLLGSGYL--- 147
Db        135 GTLGTKCIVKANHFLADL-PTKDLNQYDVTITPEVSSKSVNRAIIAELVRLYKESDLGRR 193

Qy        148 ---FDGLQL--------FTTRKFEQEIT-----VLSGKSKLDIEYKISIKFVG------- 184
Db        194 LPAYDGRKSLYTAGELPFTWKEFSVKIVDEDDGIING-PKRERSYKVAIKFVARANMHHL 252

Qy        185 --FIS-----CAEPRFLQVLNLILRR-SMKGLNLELVGRNLFDPRAKIEIREFK-MELWP 235
Db        253 GEFLAGKRADCPQ-EAVQILDIVLRELSVK--RFCPVGRSFFSPDIKTPQRLGEGLESWC 309

Qy        236 GYETSIRQHEKDI----------------------LLGTEITHKVMRTETIYDIMR--R 270
Db        310 GFYQSIRPTQMGLSLNIDMASAAFIEPLPVIEFVAQLLGKDVLSKPLSDSDRVKIKKGLR 369

Qy        271 CSHNPARHQDEVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTF----SCKGRDISFVE 326
Db        370 GVKVEVTHRANVR-------------RKYRVAGLTTQPTRELMFPVDENCTMK--SVIE 413

Qy        327 YYLTKYNIRIRDHNQPLLISKNRDKALKTNASELVVLIPELCR-VTGLNAEMRSNFQLMR 385
Db        414 YFQEMYGFTIQHTHLPCLQVGNQKKA--------SYLPMEACKIVEGQRYTKRLNEKQIT 465

Qy        386 AMSSYTRMNPK-QRTDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIV 444
Db        466 ALLKVTCQRAEGQRNDILRTVQH---NAYDQDPYAKEFGMNISEKLASVEARILPAPWLK 522

Qy        445 FH-NGK----VPAGENADWQRHFRDQRMLTTPSDGLDRWAVIAPQRNSHE--LRTLLDSL 497
Db        523 YHENGKEKDCLP--QVGQW--NMMNKKMINGMT--VSRWACVNFSRSVQENVARGFCNEL 576

Qy        498 YRAASGMGLRIRSPQEFIIYDDRTG--------TYVRAMDDCVRSDPKLILCLVPNDNAE 549
Db        577 GQMCEVSGMEFNPEPVIPIYSARPDQVEKALKHVYHTSMNKTKGKELELLLAILPDNNGS 636

Qy        550 RYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLGYTPWMI------EL 603
Db        637 LYGDLKRICETELGLISQCCLTKHVFKISKQYLA-DVSLKINVKMGGRNTVLVDAISCRI 695

Qy        604 PL-SGLMTI--GFDIA--KSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTL 658
Db        696 PLVSDIPTIIFGADVTHPENGEESSPSIAAVVASQDWPEVTKYAGLVCAQAHRQELIQDL 755

Qy        659 WP-------------MIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIE 705
Db        756 YKTWQDPVRGTVSGGMIRDLLISFRKATGQKPLRIIFYRDGVSEGQFYQVLLYEL-DAIR 814

Qy        706 KLKTEYARVQLSP---PQLAYIVVTRSMNTRFFLNGQ----------NPPPGTIVDDVIT 752
Db        815 K-----ACASLEPNYQPPVTFIVVQKRHHTRLFANNHRDKNSTDRSGNILPGTVVDTKIC 869

Qy        753 LPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPA 812
```

```
Db        870 HPTEFDFYLCSHAGIQGTSRPAHYHVLWDENNFTADGIQSLTNNLCYTYARCTRSVSIVP 929

Qy        813 VCQYA 817
Db        930 PAYYA 934
```

RESULT 12
D88568
protein ZK757.3 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C;Accession: D88568
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A;Accession: D88568
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1040 <STO>
A;Cross-references: GB:chr_III; PIDN:CAA82941.1; PID:g3877004; GSPDB:GN00021; CESP:ZK
C;Genetics:
A;Gene: ZK757.3
A;Map position: 3
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 9.8%; Score 429.5; DB 2; Length 1040;
Best Local Similarity 20.4%; Pred. No. 1.6e-21;
Matches 212; Conservative 159; Mismatches 363; Indels 303; Gaps 43;

```
Qy          9 RRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG--- 65
Db          3 RRNATSFVDNNTLTSSG-------ISGSGSMSPPITSRPASG------QASPLTSNGSLS 49

Qy         66 PPERKPWGDQYDY-----------------LNTRPVELVSKKGTDGVPVMLQTNFFRXK 107
Db         50 PPQYADDQGSVSYNLDSPRDLSPLLLSELACLNMREVVARPGLGTIGRQIPVKSNFFAMD 109

Qy        108 TK-PEWRIVHYHVEFVPTIENPRVR----------MGVLSNHANLLGSGY--LFDGL-Q 152
Db        110 LKNPKMVVIQYHVE----IHHPGCRKLDKDEMRIIFWKAVSDHPNIFHNKFALAYDGAHQ 165

Qy        153 LFTTRKFEQEITVLSGKSKLDIEYK-----------ISIKFVGFISCAEPRF------- 193
Db        166 LYTVARL--EFPDDQGSVRLDCEASLPKDNRDRTRCAISIQNVGPVLLEMQRTRTNNLDE 223

Qy        194 -----LQVLNLILRRSM----------------------KGLNLELVGRNLFDPRAKI 224
Db        224 RVLTPIQILDIICRQSLTCPLLKNSANFYTWKSSCYRIPTAAGQALDLEGGK-------- 275

Qy        225 EIREFKMELWPGYETS--IRQHEKDILLGTEITH------KVMRTETIYDIMRRCSHNPA 276
Db        276 -------EMWTGFFSSAHIASNYRP-LLNIDVAHTAFYKTRITVLQFMCDVLNERTSKPN 327

Qy        277 RH-------------------------------------QDEVRVNVLDLIV------ 291
Db        328 RNNPRGPGGPGGPGGYRGGRGGGRGGSYGNFGNRGPPGANVRDDFGGNGLTFTMDTLSRD 387

Qy        292 --LTDYNNRTY-----------------RINDVDFGQTPKSTFSCKGRD-------ISF 324
Db        388 TQLSSFETRIFGDAIRGMKIRAAHRPNAIRVYKVNSLQLPADKLMFQGIDEEGRQVVCSV 447

Qy        325 VEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASELVVLIP-ELCRVTGLNAEMRSNFQL 383
Db        448 ADYFSEKYG--------PLKYPKL--PCLHVGPPTRNIFLPMEHCLI---DSPQKYNKKM 494

Qy        384 MRAMSS----YTRMNPKQRTDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIG 439
Db        495 SEKQTSAIIKAAAVDATQREDRIKQL--AAQASFGTDPFLKEFGVAVSSQMIQTTARVIQ 552
```

```
Qy        440 QQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDG---LDRWAVIAP------------ 483
                 |:|       | |       |    :  | ||   :|   :  |
Db        553 PPPIMF------GGNN-------RSVNPVVFPKDGSWTMDNQTLYMPATCRSYSMIALVD 599

Qy        484 QRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIY----DDRTGTYVRAMD---------D 530
               |:   |:|   ||   |: ||:      : : |    :|    :    |          |
Db        600 PRDQTSLQTFCQSLTMKATAMGMNFPRWPDLVKYGRSKEDVCTLFTEIADEYRVTNTVCD 659

Qy        531 CVRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQL 590
               |       |: :: : |:: | ::|::   :   : :| | :|   :|   :    | ::|
Db        660 C-------IIVVLQSKNSDIYMTVKEQSDIVHGIMSQCVLMKNV-SRPTPATCANIVLKL 711

Qy        591 NCKLG-----------YTPWMIELPLSGLMTIGFDIAKSTRDRKR----AYGALIASMDL 635
               | |:|              :::: |    | :| |:   |:   |    :  |::|::||
Db        712 NMKMGGINSRIVADKITNKYLVDQP---TMVVGIDVTHPTQAEMRMNMPSVAAIVANVDL 768

Qy        636 QQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQL 695
                    |   :    : :   |  :  : :  |:|  :| |:||: |||||| |   ::
Db        769 LPQSYGANVKVQKKCRESVVYLLDAIRERIITFYRHTKQK-PARIIVYRDGVSEGQFSEV 827

Qy        696 FEFEVKDIIEKLKTEYARV--QLSPPQLAYIVVTRSMNTRFFLNGQNP--------PPGT 745
                  ::: |: ::|    :     || : |||| :  : | |   ||          ||||
Db        828 ----LREEIQSIRTACLAIAEDFRPP-ITYIVVQKRHHARIFCKYQNDMVGKAKNVPPGT 882

Qy        746 IVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWS 805
               ||  | || :|||| |    ||| |  |:||      : :::| :|| ||| |   :
Db        883 TVDTGIVSPEGFDFYLCSHYGVQGTSRPARYHVLLDECKFTADEIQSITYGMCHTYGRCT 942

Qy        806 GTTRVPAVCQYAKKLAT 822
               :  :|    ||  :||
Db        943 RSVSIPTPVYYADLVAT 959
```

RESULT 13
S41013
hypothetical protein ZK757.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Feb-2000
C;Accession: S41013
R;Thomas, K.
submitted to the EMBL Data Library, December 1993
A;Reference number: S41011
A;Accession: S41013
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-958 <THO>
A;Cross-references: EMBL:Z29121
C;Genetics:
A;Introns: 52/2; 350/2; 422/3; 592/1; 679/3; 721/2; 887/2; 929/1
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 9.6%; Score 420.5; DB 2; Length 958;
Best Local Similarity 20.6%; Pred. No. 6.1e-21;
Matches 194; Conservative 155; Mismatches 322; Indels 271; Gaps 40;

```
Qy         85 ELVSKK--GTDGVPVMLQTNFFRXKTK-PEWRIVHYHVEFVPTIENPRVR---------- 131
              |:|::   || |  : :::|||    | |:  :: ||||    | :|  |
Db          3 EVVARPGLGTIGRQIPVKSNFFAMDLKNPKMVVIQYHVE----IHHPGCRKLDKDEMRII 58

Qy        132 -MGVLSNHANLLGSGY--LFDGL-QLFTTRKFEQEITVLSGKSKLDIEYK---------- 177
                 :|:| |:  : :   :|| ||:|  :   |     |  :|| |
Db         59 FWKAVSDHPNIFHNKFALAYDGAHQLYTVARL--EFPDDQGSVRLDCEASLPKDNRDRTR 116

Qy        178 --ISIKFVGFISCAEPRF-----------LQVLNLILRRSM------------------ 205
                |||: || :     |             :|:|::| |:|:
Db        117 CAISIQNVGPVLLEMQRTRTNNLDERVLTPIQILDIICRQSLTCPLLKNSANFYTWKSSC 176

Qy        206 ------KGLNLELVGRNLFDPRAKIEIREFKMELWPGYETS--IRQHEKDILLGTEITH- 256
                    |  |:| |                 |:| |: :|  |  : :  ||  :: |
Db        177 YRIPTAAGQALDLEGGK--------------EMWTGFFSSAHIASNYRP-LLNIDVAHT 220

Qy        257 -----KVMRTETIYDIMRRCSHNPARH---------------------------------- 278
```

```
                  ::   : : |::   :  | |:
Db        221 AFYKTRITVLQFMCDVLNERTSKPNRNNPRGPGGPGGPGGYRGGRGGGRGGSYGNFGNRG 280

Qy        279 ------QDEVRVNVLDLIV--------LTDYNNRTY-----------------RINDVD 306
                   :|:   | |   :         |: :  | :                 |:  |:
Db        281 PPGANVRDDFGGNGLTFTMDTLSRDTQLSSFETRIFGDAIRGMKIRAAHRPNAIRVYKVN 340

Qy        307 FGQTPKSTFSCKGRD-------ISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASE 359
               | |      :| |       |  :|:  ||          ||   |     |
Db        341 SLQLPADKLMFQGIDEEGRQVVCSVADYFSEKYG--------PLKYPKL--PCLHVGPPT 390

Qy        360 LVVLIP-ELCRVTGLNAEMRSNFQLMRAMSS----YTRMNPKQRTDRLRAFNHRLQNTPE 414
               : :| | | :   ::  : | ::    :|       ::  || ||::      | :
Db        391 RNIFLPMEHCLI---DSPQKYNKKMSEKQTSAIIKAAAVDATQREDRIKQL--AAQASFG 445

Qy        415 SVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDG 474
               :   |::: : :   : :   |:|    |:|       | |       |    :  | ||
Db        446 TDPFLKEFGVAVSSQMIQTTARVIQPPPIMF------GGNN-------RSVNPVVFPKDG 492

Qy        475 ---LDRWAVIAP-------------QRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIY- 517
                  :|   :  |             |:   |:|   ||   |: ||:      : : |
Db        493 SWTMDNQTLYMPATCRSYSMIALVDPRDQTSLQTFCQSLTMKATAMGMNFPRWPDLVKYG 552

Qy        518 ---DDRTGTYVRAMD---------DCVRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVP 565
                  :|    :    |         ||       |: :: : |:: | ::|::   :   :
Db        553 RSKEDVCTLFTEIADEYRVTNTVCDC-------IIVVLQSKNSDIYMTVKEQSDIVHGIM 605

Qy        566 TQVVTLKTTKNRSLMSIATKIAIQLNCKLG-----------YTPWMIELPLSGLMTIGFD 614
               :| | :|   :|   :    | ::|| |:|             :::: |    | :| |
Db        606 SQCVLMKNV-SRPTPATCANIVLKLNMKMGGINSRIVADKITNKYLVDQP---TMVVGID 661

Qy        615 IAKSTRDRKR----AYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQ 670
               :   |:   |    :  |::|::||  |   :   :   :  |  : : |:  |:
Db        662 VTHPTQAEMRMNMPSVAAIVANVDLLPQSYGANVKVQKKCRESVVYLLDAIRERIITFYR 721

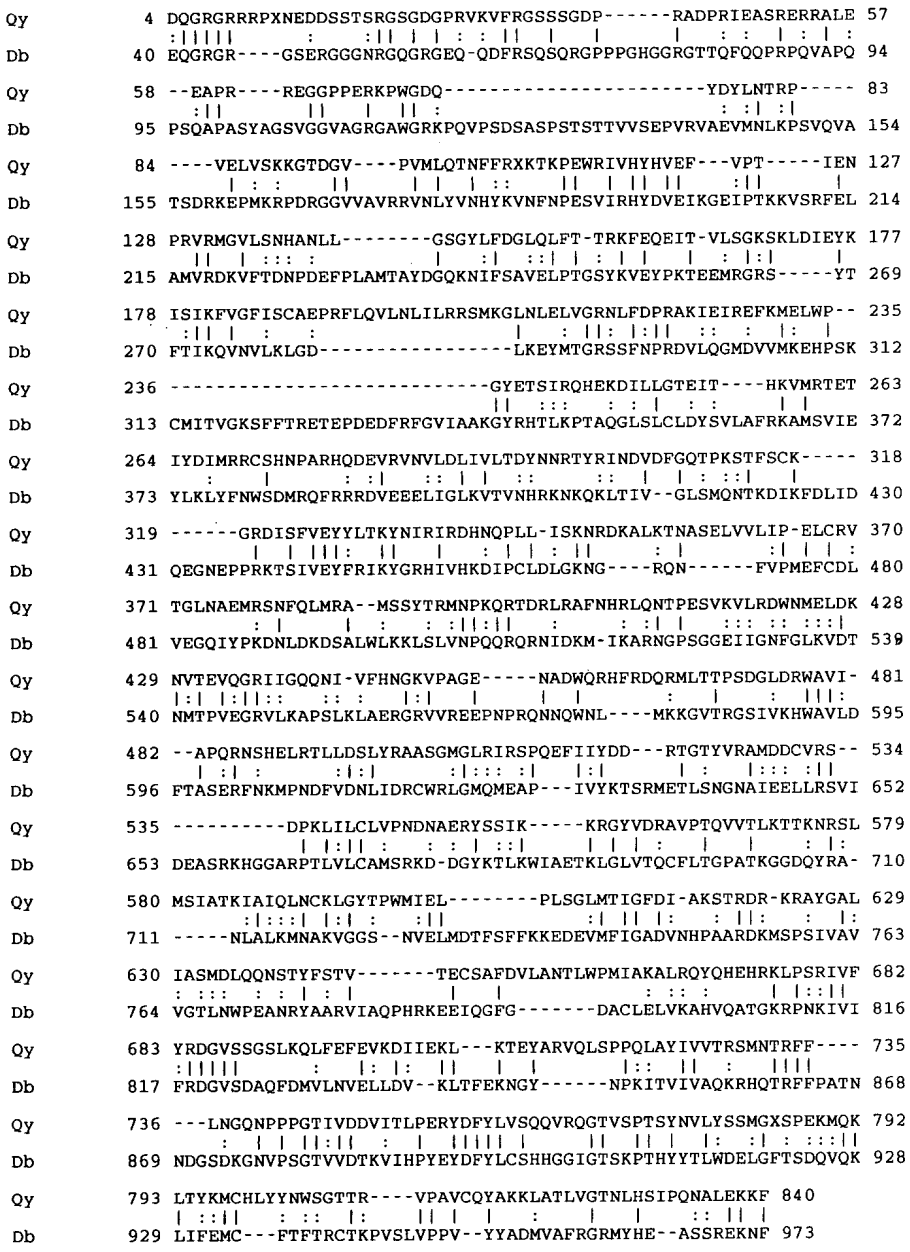
Qy        671 HEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARV--QLSPPQLAYIVVTR 728
               |  :| |:||: |||||| |   ::    ::: |: ::|    :      || : |||| :
Db        722 HTKQK-PARIIVYRDGVSEGQFSEV----LREEIQSIRTACLAIAEDFRPP-ITYIVVQK 775

Qy        729 SMNTRFFLNGQNP--------PPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLY 780
               : | |   ||          |||| ||  |  || :|||| |    |||  |  |:||
Db        776 RHHARIFCKYQNDMVGKAKNVPPGTTVDTGIVSPEGFDFYLCSHYGVQGTSRPARYHVLL 835

Qy        781 SSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLAT 822
                  : :::| :|| ||| |   : :  :|    ||  :||
Db        836 DECKFTADEIQSITYGMCHTYGRCTRSVSIPTPVYYADLVAT 877
```

RESULT 14
H86438
protein T19E23.7 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: H86438
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H86438
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1014 <STO>
A;Cross-references: GB:AE005172; NID:g6692120; PIDN:AAF24585.1; GSPDB:GN00141
C;Genetics:
A;Gene: T19E23.7
A;Map position: 1

Query Match 8.2%; Score 360.5; DB 2; Length 1014;
Best Local Similarity 21.2%; Pred. No. 1e-16;
Matches 215; Conservative 166; Mismatches 378; Indels 253; Gaps 49;



RESULT 15
T01113
translation initiation factor eIF-2C homolog T21L14.12 - Arabidopsis thaliana
N;Alternate names: Argonaute (AGO1)-like protein
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 16-Feb-2001
C;Accession: T01113; D84739
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K
submitted to the EMBL Data Library, December 1997
A;Description: Arabidopsis thaliana chromosome II BAC T21L14 genomic sequence.
A;Reference number: Z14209
A;Accession: T01113
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-887 <ROU>
A;Cross-references: EMBL:AC003033; NID:g2702261; PID:g2702284
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: D84739
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-887 <STO>
A;Cross-references: GB:AE002093; NID:g2702284; PIDN:AAB91987.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g32940; T21L14.12
A;Map position: 2
A;Introns: 61/3; 130/2; 203/2; 248/1; 281/3; 314/2; 367/3; 406/3; 442/3; 463/3; 501/3
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 8.1%; Score 353.5; DB 2; Length 887;
Best Local Similarity 21.3%; Pred. No. 2.5e-16;
Matches 195; Conservative 135; Mismatches 336; Indels 251; Gaps 41;



```
Db        448 NEDFQPCNGRWNFNNKMLLEPR-AIKSWAIVNFSFPCDSSHISRELI------SCGMRKG 500

Qy        506 LRIRSPQEFIIYD---------DRTGTYVRAMDDCVRSDPKLILCLVP-NDNAERYSSIK 555

Db        501 IEIDRPFALVEEDPQYKKAGPVERVEKMIATMKLKFPDPPHFILCILPERKTSDIYGPWK 560

Qy        556 KRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDI 615

Db        561 KICLTEEGIHTQCICPIKISDQYL----TNVLLKINSKLG--------GINSLLGIEY-- 606

Qy        616 AKSTRDRKRAYG--------ALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAK--- 664

Db        607 ---------SYNIPLINKIPTLILGMDVSHGPPGRADVPSVAA--VVGSKCWPLISRYRA 655

Qy        665 ALRQ---------------------------------YQHEHRKLPSRIVFYRDGVSSGS 691

Db        656 AVRTQSPRLEMIDSLFQPIENTEKGDNGIMNELFVEFYRTSRARKPKQIIIFRDGVSESQ 715

Qy        692 LKQLFEFEVKDIIEKLKTEYARV-QLSPPQLAYIVVTRSMNTRFF--LNGQNPPPGTIVD 748

Db        716 FEQVLKIEVDQIIK----AYQRLGESDVPKFTVIVAQKNHHTKLFQAKGPENVPAGTVVD 771

Qy        749 DVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQ----KLTYKMCHLYYNW 804

Db        772 TKIVHPTNYDFYMCAHAGKIGTSRPAHYHVLLDEIGFSPDDLQNLIHSLSYKLLNSIFNV 831

Qy        805 SGTTRVPAVCQYAKKLA 821

Db        832 SS-----LLCVFVLSVA 843
```

Search completed: July 22, 2003, 07:05:49
Job time : 20.7549 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2003, 22:53:12 ; Search time 9.85581 Seconds
(without alignments)
3547.609 Million cell updates/sec

Title: US-09-873-737A-2
Perfect score: 4389
Sequence: 1 MADDQGRGRRRPXNEDDSST.....VGTNLHSIPQNALEKKFYYL 843

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	478.5	10.9	857	1	I2C1_HUMAN	Q9ul18 homo sapien
2	472	10.8	863	1	I2C2_RAT	Q9qz81 rattus norv
3	454.5	10.4	813	1	I2C2_RABIT	O77503 oryctolagus
4	447	10.2	834	1	IF2C_SCHPO	O74957 schizosacch
5	436.5	9.9	988	1	PINH_ARATH	Q9xgw1 arabidopsis
6	433	9.9	997	1	AGOL_ARATH	Q9sjk3 arabidopsis
7	431	9.8	1048	1	AGO1_ARATH	O04379 arabidopsis
8	429.5	9.8	1040	1	YO43_CAEEL	P34681 caenorhabdi
9	321	7.3	377	1	I2C2_HUMAN	Q9ukv8 homo sapien
10	256.5	5.8	1032	1	YQ53_CAEEL	Q09249 caenorhabdi
11	129.5	3.0	882	1	RA50_PYRFU	P58301 pyrococcus
12	124.5	2.8	5147	1	FAT_DROME	P33450 drosophila
13	121.5	2.8	2022	1	ANT1_ONCVO	P21249 onchocerca
14	116.5	2.7	1287	1	SKI2_YEAST	P35207 saccharomyc
15	115.5	2.6	943	1	ARS2_DROME	Q9v9k7 drosophila
16	115.5	2.6	1803	1	YJL3_YEAST	P47024 saccharomyc
17	114	2.6	1021	1	SECA_ARATH	Q9syi0 arabidopsis
18	112	2.6	879	1	MYSP_DROME	P35415 drosophila
19	111.5	2.5	2349	1	TPR_HUMAN	P12270 homo sapien
20	111	2.5	1702	1	DPOL_THELI	P30317 thermococcu
21	110.5	2.5	870	1	CSX2_SCHPO	Q9uue2 schizosacch
22	110	2.5	750	1	CBBB_BACTV	Q9ziu5 bacillus th
23	110	2.5	958	1	YBS3_YEAST	P38086 saccharomyc
24	110	2.5	1025	1	MK21_YEAST	Q12176 saccharomyc
25	109	2.5	950	1	Y511_RICPR	Q9zd36 rickettsia
26	108.5	2.5	675	1	DNLJ_BUCAI	P57172 buchnera ap
27	108.5	2.5	861	1	GCR3_YEAST	P34160 saccharomyc
28	108	2.5	1727	1	ALM1_SCHPO	Q9utk5 schizosacch
29	108	2.5	1845	1	YM68_CAEEL	P34529 caenorhabdi
30	108	2.5	1919	1	HAPP_RAT	P97924 rattus norv
31	108	2.5	2748	1	NUM1_YEAST	Q00402 saccharomyc
32	107.5	2.4	1804	1	YFA7_YEAST	P43583 saccharomyc
33	107	2.4	2077	1	TEGU_HSV6G	P30002 human herpe
34	106.5	2.4	813	1	T1RE_ECOLI	Q47281 escherichia
35	106	2.4	782	1	MUS2_STAAM	Q99uw1 staphylococ
36	106	2.4	910	1	SC15_YEAST	P22224 saccharomyc
37	106	2.4	1435	1	LTE1_YEAST	P07866 saccharomyc
38	105.5	2.4	1265	1	CYA5_CANFA	P30803 canis famil
39	105.5	2.4	1509	1	MYSN_ACACA	P05659 acanthamoeb
40	105.5	2.4	1839	1	CYAA_SACKL	P23466 saccharomyc
41	105	2.4	2245	1	MYSJ_DICDI	P54697 dictyosteli
42	104.5	2.4	378	1	YIBH_ECOLI	P32107 escherichia
43	104.5	2.4	850	1	STB2_YEAST	P46679 saccharomyc
44	104	2.4	414	1	YJL4_YEAST	P47023 saccharomyc
45	104	2.4	1336	1	SEC3_YEAST	P33332 saccharomyc

ALIGNMENTS

RESULT 1
I2C1_HUMAN
ID I2C1_HUMAN STANDARD; PRT; 857 AA.
AC Q9UL18;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Eukaryotic translation initiation factor 2C 1 (eIF2C 1) (eIF-2C 1)
DE (Putative RNA-binding protein Q99).
GN EIF2C1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20005943; PubMed=10534406;
RA Koesters R., Adams V., Betts D., Moos R., Schmid M., Siermann A.,
RA Hassam S., Weitz S., Lichter P., Heitz P.U., von Knebel Doeberitz M.,
RA Briner J.;
RT "Human eukaryotic initiation factor EIF2C1 gene: cDNA sequence,
RT genomic organization, localization to chromosomal bands 1p34-p35, and
RT expression.";
RL Genomics 61:210-218(1999).
CC -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC initiation process.
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; AF093097; AAF00068.1; -.
DR Genew; HGNC:3262; EIF2C1.
DR MIM; 606228; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
KW Initiation factor; Protein biosynthesis.
FT DOMAIN 226 346 PAZ.
FT DOMAIN 515 816 PIWI.
SQ SEQUENCE 857 AA; 97214 MW; 1DBB524AE7CBAF66 CRC64;

Query Match 10.9%; Score 478.5; DB 1; Length 857;
Best Local Similarity 22.7%; Pred. No. 5e-26;
Matches 197; Conservative 148; Mismatches 338; Indels 185; Gaps 35;

```
Qy         58 EAPRREGGPPERKPWGDQYDYLNTRPVELVSKKGTDGVPVMLQTNFFRXKTKPEWRIVHY 117
              :|||| |                                 || | |: |  |:|     |:  : ||
Db         22 QAPRRPG------------------------IGTVGKPIKLLANYFEVDI-PKIDVYHY 55

Qy        118 HVEFVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLS-GKSKLDIE- 175
               |:  |     ||   |:             |  ::  :|    :| |  |  ::| |
Db         56 EVDIKPDKCPRRVNREVVEYMVQHFKPQIFGDRKPVYDGKKNIYTVTALPIGNERVDFEV 115

Qy        176 ----------YKISIKFVGFISCAEPRFL---------------QVLNLILRRSMKGLN'209
                    :|:|||::  :|    | |                 | |:: :|  :  :
Db        116 TIPGEGKDRIFKVSIKWLAIVSW---RMLHEALVSGQIPVPLESVQALDVAMRH-LASMR 171

Qy        210 LELVGRNLFDP-----------RAKIEIREFKMELWPGYETSIRQHEKDILLGTEITHKV 258
               |||: | |            |          |:| |:  |:|     ::|  :::
Db        172 YTPVGRSFFSPPEGYYHPLGGGR----------EVWFGFHQSVRPAMWKMMLNIDVSATA 221

Qy        259 M-RTETIYDIM------RRCSHNPARHQDEVRV----NVLDLIVLTDY---NNRTYRIND 304
               : : : : |       |      |    | ||      : | |   :      | ||: :
Db        222 FYKAQPVIEFMCEVLDIRNIDEQPKPLTDSQRVRFTKEIKGLKVEVTHCGQMKRKYRVCN 281

Qy        305 VDFGQTPKSTF-----SCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASE 359
               |         ||    | :  : :  :|:  |||:::  :  |             |:    :
Db        282 VTRRPASHQTFPLQLESGQTVECTVAQYFKQKYNLQLKYPHLP---------CLQVGQEQ 332

Qy        360 LVVLIP-ELCRVTGLNAEMRSNFQLM-RAMSSYTRMNPKQRTDRLRAFNHRLQNTPESV- 416
                :| |:| :    |  |   :|     |:  :   :   ||    :  ::|   ::
Db        333 KHTYLPLEVCNIV---AGQRCIKKLTDNQTSTMIKATARSAPDRQEEISRLMKNASYNLD 389

Qy        417 KVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDG-- 474
               :::: :::  ::||| ||::    |: : |:                | : ||:.|
Db        390 PYIQEFGIKVKDDMTEVTGRVL-PAPILQYGGR--------------NRAIATPNQGVW 433

Qy        475 ------------LDRWAV--IAPQRNSHE--LRTLLDSLYRAASGMGLRIRSPQEFIIY- 517
                        :  ||:   |||:   |  |:   | | : :   |: |:    |  |
Db        434 DMRGKQFYNGIEIKVWAIACFAPQKQCREEVLKNFTDQLRKISKDAGMPIQGQPCFCKYA 493

Qy        518 --DDRTGTYVRAMDDCVRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTK 575
                 |      | : :   |  :||: ::|      |: :|: |     : || | :|
Db        494 QGADSVEPMFRHLKN-TYSGLQLIIVILPG-KTPVYAEVKRVGDTLLGMATQCVQVKNVV 551

Qy        576 NRSLMSIATKIAIQLNCKLG---------YTPWMIELPLSGLMTIGFDIA--KSTRDRKR 624
               |  :: : :::| |||             : : |   :: :| |:    :   :|
Db        552 KTSPQTL-SNLCLKINVKLGGINNILVPHQRSAVFQQP---VIFLGADVTHPPAGDGKKP 607

Qy        625 AYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYR 684
               :  |:: |||    | | :||        :   |  |: : | |:    |  |:||:|||
Db        608 SITAVVGSMDAHP-SRYCATVRVQRPRQEIIEDLSYMVRELLIQFYKSTRFKPTRIIFYR 666

Qy        685 DGVSSGSLKQLFEFE---VKDIIEKLKTEYARVQLSPPQLAYIVVTRSMNTRFFLNGQNP 741
               |||  | | |:  :|   ::|   ||: :|       | : ||||  :  :|| |   :|
Db        667 DGVPEGQLPQILHYELLAIRDACIKLEKDY------QPGITYIVVQKRHHTRLFCADKNE 720

Qy        742 --------PPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKL 793
                    | || ||  || |  :|||| |    |||  |: | ||:     : :::| |
Db        721 RIGKSGNIPAGTTVDTNITHPFEFDFYLCSHAGIQGTSRPSHYYVLWDDNRFTADELQIL 780

Qy        794 TYKMCHLYYNWSGTTRVPAVCQYAKKLA 821
               ||::|| |   : :  :||  ||: :|
Db        781 TYQLCHTYVRCTRSVSIPAPAYYARLVA 808


RESULT 2
I2C2_RAT
ID   I2C2_RAT       STANDARD;      PRT;   863 AA.
AC   Q9QZ81;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2)
DE   (Golgi ER protein 95 kDa) (GERp95).
GN   EIF2C2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Hepatoma;
RX   MEDLINE=99443791; PubMed=10512872;
RA   Cikaluk D.E., Tahbaz N., Hendricks L.C., DiMattia G.E., Hansen D.,
RA   Pilgrim D., Hobman T.C.;
RT   "GERp95, a membrane-associated protein that belongs to a family of
RT   proteins involved in stem cell differentiation.";
RL   Mol. Biol. Cell 10:3357-3372(1999).
CC   -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC       initiation process.
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF195534; AAF12800.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Initiation factor; Protein biosynthesis.
FT   DOMAIN      239    352       PAZ.
FT   DOMAIN      521    822       PIWI.
SQ   SEQUENCE   863 AA;  97659 MW;  9A1E0D6B84F7B622 CRC64;

  Query Match           10.8%;  Score 472;  DB 1;  Length 863;
  Best Local Similarity  23.0%;  Pred. No. 1.4e-25;
  Matches  200;  Conservative  142;  Mismatches  300;  Indels  226;  Gaps   39;

Qy         91 GTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMGVLSN-----HANLLGSG 145
              || |  : || |||     |:  | || ::  |     ||   :: :         : |
Db         36 GTTGRTIKLQANFFEMDI-PKIDIYHYELDIKPEKCPRRVNREIVEHMVQHFKTQIFGDR 94

Qy        146 Y-LFDGLQ-LFTTR-------KFEQEITVLSGKSKLDIEYKISIKFVGFISC-------- 188
               :||| : |:|          | | |:| | | |: | |  :|:|||:|  :|
Db         95 KPVFDGRKNLYTAMPLPIGRDKVELEVT-LPGEGK-DRIFKVSIKWVSCVSLQALHDALS 152

Qy        189 ----AEP-RFLQVLNLILRRSMKGLNLELVGRNLF-------DPRAKIEIREFKMELWPG 236
                  : |   :| |::::|  :  :    |||: |       :|           |:| |
Db        153 GRLPSVPFETIQALDVVMRH-LPSMRYTPVGRSFFTASEGCSNPLGG------GREVWFG 205

Qy        237 YETSIR-----------------------------------QHEKDIL---------- 249
               :  |:|                                  : :| :
Db        206 FHQSVRPSLWKMMLNIDVSATAFYKAQPVIEFVCEVLDFKSIEEQQKPLTDSQRVKFTKE 265

Qy        250 ---LGTEITH--KVMRTETIYDIMRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTYRIND 304
                  |  ||||  :: |   : :: ||    || ||                    |: :
Db        266 IKGLKVEITHCGQMKRKYRVCNVTRR----PASHQ------------------TFPLQQ 302

Qy        305 VDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASELVVLI 364
               : ||| : | :          :|:  :: : :|  : |          |:    :   :
Db        303 -ESGQTVECTVA---------QYFKDRHKLVLRYPHLP---------CLQVGQEQKHTYL 343

Qy        365 P-ELCRVTG-------LNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTPESV 416
               | |:| :          |       |       | |:  |       :  |  :||      :
Db        344 PLEVCNIVAGQRCIKKLTDNQTSTMIRATARSAPDRQEEISKLMRSASFN--------TD 395

Qy        417 KVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDG-- 474
               :|:: : :   :|:| ||::   :|::       | |          : : ||  |
Db        396 PYVREFGIMVKDEMTDVTGRVLQPPSILY------GGRN----------KAIATPVQGVW 439
```

```
Qy        475 -----------LDRWAV--IAPQRNSHE--LRTLLDSLYRAASGMGLRIRSPQEFIIY- 517
                          :  ||:   ||||   |  |::  : | : :   |: |:    |  |
Db        440 DMRNKQFHTGIEIKVWAIACFAPQRQCTEVHLKSFTEQLRKISRDAGMPIQGQPCFCKYA 499

Qy        518 --DDRTGTYVRAMDDCVRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTK 575
                 |      | : :   :  :|:: ::|      |: :|: |     : || | :|  :
Db        500 QGADSVEPMFRHLKN-TYAGLQLVVVILPG-KTPVYAEVKRVGDTVLGMATQCVQMKNVQ 557

Qy        576 NRSLMSIATKIAIQLNCKLGYT---------PWMIELPLSGLMTIGFDIA--KSTRDRKR 624
                |:     : : :::| |||           | : : |   :: :| |:    :   :|
Db        558 -RTTPQTLSNLCLKINVKLGGVNNILLPQGRPPVFQQP---VIFLGADVTHPPAGDGKKP 613

Qy        625 AYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYR 684
                :  |:: |||   |  | :||       :   |  |: : | |:    |  |:||:|||
Db        614 SIAAVVGSMDAHPN-RYCATVRVQQHRQEIIQDLAAMVRELLIQFYKSTRFKPTRIIFYR 672

Qy        685 DGVSSGSLKQLFEFEVKDIIE---KLKTEYARVQLSPPQLAYIVVTRSMNTRFFLNGQNP 741
                |||| |  :|:   |:  | |   ||: ||       | : :||| :  :|| |   :|
Db        673 DGVSEGQFQQVLHHELLAIREACIKLEKEY------QPGITFIVVQKRHHTRLFCTDKNE 726

Qy        742 --------PPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKL 793
                      | || ||  || |  :|||| |    |||  |: |:|||:     | :::| |
Db        727 RVGKSGNIPAGTTVDTKITHPTEFDFYLCSHAGIQGTSRPSHYHVLWDDNRFSSDELQIL 786

Qy        794 TYKMCHLYYNWSGTTRVPAVCQYAKKLA 821
                ||::|| |   :  :  :||   ||  :|
Db        787 TYQLCHTYVRCTRSVSIPAPAYYAHLVA 814
```

RESULT 3
I2C2_RABIT
ID I2C2_RABIT STANDARD; PRT; 813 AA.
AC O77503;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2).
GN EIF2C2.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=98267198; PubMed=9602122;
RA Zou C., Zhang Z., Wu S., Osterman J.C.;
RT "Molecular cloning and characterization of a rabbit eIF2C protein.";
RL Gene 211:187-194(1998).
CC -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC initiation process.
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; AF005355; AAC24323.1; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
KW Initiation factor; Protein biosynthesis.
FT DOMAIN 189 302 PAZ.
FT DOMAIN 471 772 PIWI.

SQ SEQUENCE 813 AA; 92281 MW; 13BA0DB70E10688F CRC64;

Query Match 10.4%; Score 454.5; DB 1; Length 813;
Best Local Similarity 22.7%; Pred. No. 2.3e-24;
Matches 193; Conservative 142; Mismatches 289; Indels 225; Gaps 38;

```
Qy        110 PEWRIVHYHVEFVPTIENPRVRMGVLSN-----HANLLGSGY-LFDGLQ-LFTT------ 156
              |:  | || ::  |     ||   :: :      | : |    :||| : |:|
Db          4 PKIDIYHYELDIKPEKCPRRVNREIVEHMVQHFKAQIFGDRKPVFDGRKNLYTAMPLPIG 63

Qy        157 -RKFEQEITVLSGKSKLDIEYKISIKFVGFISC-----------AEP-RFLQVLNLILR 202
               | | |:| | |: | |  :|:|||:|  :|             : |   :| |::::|
Db         64 REKVELEVT-LPGEGK-DRIFKVSIKWVSCVSLQALHDALSGRLPSVPFETIQALDVVMR 121

Qy        203 RSMKGLNLELVGRNLF-------DPRAKIEIREFKMELWPGYETSIR------------ 242
                 :    |||: |         :|          |:| |:  |:|
Db        122 H-LPSMRYTPVGRSFFTASEGCSNPLGG------GREVWFGFHQSVRPSLWKMMLNIDVS 174

Qy        243 ----------------------QHEKDIL-------------LGTEITH--KVMRTET 263
                                     : :| :                |  |||| :: |
Db        175 ATAFYKAQPVIEFVCEVLDFKSIEEQQKPLTDSQRVKFTKEIKGLKVEITHCGQMKRKYR 234

Qy        264 IYDIMRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTFSCKGRDIS 323
              : :: ||    || ||                    |: :   : |||  : | :
Db        235 VCNVTRR----PASHQ-------------------TFPLQQ-ESGQTVECTVA------- 263

Qy        324 FVEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASELVVLIP-ELCRVTG-------LNA 375
                :|:  :: : :|  : |          |:    :    :| |:| :          |
Db        264 --QYFKDRHKLVLRYPHLP---------CLQVGQEQKHTYLPLEVCNIVAGQRCIKKLTD 312

Qy        376 EMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQG 435
                |      | |:  |      :  |  :||         :   :|:: : :   :|:| |
Db        313 NQTSTMIRATARSAPDRQEEISKLMRSASFN--------TDPYVREFGIMVKDEMTDVTG 364

Qy        436 RIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDG-------------LDRWAV- 480
                |::   :|::       | |          : : ||  |              :  ||:
Db        365 RVLQPPSILY------GGRN----------KAIATPVQGVWDMRNKQFHTGIEIKVWAIA 408

Qy        481 -IAPQRNSHE--LRTLLDSLYRAASGMGLRIRSPQEFIIY---DDRTGTYVRAMDDCVRS 534
               ||||   |  |::  : | : :   |: |:    |  |    |  | : :   :
Db        409 CFAPQRQCTEVHLKSFTEQLRKISRDAGMPIQGQPCFCKYAQGADSVGPMFRHLKN-TYA 467

Qy        535 DPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKL 594
               :|:: ::|      |: :|: |     : || | :|  : |:     : : :::| ||
Db        468 GLQLVVVILPG-KTPVYAEVKRVGDTVLGMATQCVQMKNVQ-RTTPQTLSNLCLKINVKL 525

Qy        595 GYT---------PWMIELPLSGLMTIGFDIA--KSTRDRKRAYGALIASMDLQQNSTYFS 643
               |          | : : |   :: :| |:    :   :| :  |:: |||   |  | :
Db        526 GGVNNILLPQGRPPVFQQP---VIFLGADVTHPPAGDGKKPSIAAVVGSMDAHPN-RYCA 581

Qy        644 TVTECSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDI 703
               ||        :   |  |: : | |:    |  |:||:||||||| |  :|:   |: |
Db        582 TVRVQQHRQEIIQDLAAMVRELLIQFYKSTRFKPTRIIFYRDGVSEGQFQQVLHHELLAI 641

Qy        704 IE---KLKTEYARVQLSPPQLAYIVVTRSMNTRFFLNGQNP--------PPGTIVDDVIT 752
               |   ||: :|       | : :||| :  :|| |   :|         | || ||  ||
Db        642 REACIKLEKDY------QPGITFIVVQKRHHTRLFCTDKNERVGKSGNIPAGTTVDTKIT 695

Qy        753 LPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPA 812
               |  :|||| |    |||  |: |:|||:     | :::| |||::|| |   : :  :||
Db        696 HPTEFDFYLCSHAGIQGTSRPSHYHVLWDDNRFSSDELQILTYQLCHTYVRCTRSVSIPA 755

Qy        813 VCQYAKKLA 821
                ||  :|
Db        756 PAYYAHLVA 764
```

RESULT 4
IF2C_SCHPO
ID IF2C_SCHPO STANDARD; PRT; 834 AA.
AC O74957;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable eukaryotic translation initiation factor 2C (eIF2C) (eIF-2C).
GN   SPCC736.11.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC       initiation process (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AL023705; CAA19275.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Initiation factor; Protein biosynthesis.
FT   DOMAIN      212    326       PAZ.
FT   DOMAIN      500    799       PIWI.
SQ   SEQUENCE   834 AA;  94438 MW;  F5CE397BF590D9E4 CRC64;

  Query Match           10.2%;  Score 447;  DB 1;  Length 834;
  Best Local Similarity  21.5%;  Pred. No. 8.1e-24;
  Matches  194;  Conservative  157;  Mismatches  332;  Indels  220;  Gaps   37;

Qy         85 ELVSKKGTDGV--PVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMGVLSNHANL- 141
              |:  : |  |:   : |: |||:  : |   |  |||        || :  :: |   :
Db          8 EIALRPGYGGLGKQITLKANFFQIISLPNETINQYHVIVGDGSRVPRKQSQLIWNSKEVK 67

Qy        142 --LGSGYL---FDGLQLFTTRKFEQEITVLSGKSKLDI-------EYKISI--------- 180
                 || ::   :||  :  ::       :  |  |::|       | : ||
Db         68 QYFGSSWMNSVYDGRSMCWSKG-----DIADGTIKVNIGSESHPREIEFSIQKSSKINLH 122

Qy        181 ---KFVGFISCAEPRFL---QVLNLILRRS-----------MKGLNLELVGRNLFDPRA 222
                :||     ::|: |     |:|:|::             | |   :|
Db        123 TLSQFVNSKYSSDPQVLSSIMFLDLLLKKKPSETLFGFMHSFFTGENGVSLGGG------ 176

Qy        223 KIEIREFKMELWPGYETSIRQHEKDILLGTEITHKVM-RTETIYDIMRR---CSHNPARH 278
                :| | |:  ||| ::  : :  :|:     | :::  |:     ||
Db        177 --------VEAWKGFYQSIRPNQGFMSVNVDISSSAFWRNDSLLQILMEYTDCS------ 222

Qy        279 QDEVRVNVLDLI----------------------VLTDYNNRTYRINDVDFGQTPKSTF 315
                || ||                          | ||  || | |          | |
Db        223 ------NVRDLTRFDLKRLSRKFRFLKVTCQHRNNVGTDLANRVYSIEGFSSKSASDSFF 276

Qy        316 ----SCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASELVVLIP-ELCRV 370
                  : : ||  ||:|  :|:|::  | | :: ||             ::| | | |
Db        277 VRRLNGEEQKISVAEYFLENHNVRLQYPNLPCILVKNG----------AMLPIEFCFV 324

Qy        371 -------TGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLRDWN 423
                    ||::  :|     |  :   | :|  ::  | |::   ::   |  :
Db        325 VKGQRYTAKLNSDQTAN------MIRFAVQRPFERVQQIDDFVHQMD--WDTDPYLTQYG 376

Qy        424 MELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDGLDRWAVIA- 482
              |:: | : ||  |::   :| :    :    :  | : | :| |  |   :  |||:
Db        377 MKIQKKMLEVPARVLETPSIRYGGDCIERPVSGRW--NLRGKRFLDPPRAPIRSWAVMCF 434

Qy        483 ------PQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTG-------TYVRAMD 529
                   | |    :  | :  :  : :|:     :  ::| | |  |       | :  :
Db        435 TSTRRLPMRG---IENFLQTYVQTLTSLGINFVMKKPPVLYADIRGSVEELCITLYKKAE 491

Qy        530 DCVRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLK---TTKNRSLMSIATKI 586
              : |  :  ::  :: | | |||:       ||:|    |    :| :   ::  ||
Db        492 QVGNAPPDYLFFILDKNSPEPYGSIKRVCNTMLGVPSQCAISKHILQSKPQYCANLGMKI 551

Qy        587 AIQ---LNCK-------LGYTPWMI---ELPLSGLMTIGFDIAKSTRDRKRAYGALIASM 633
              ::   :||        ||  | :|  ::  |:   |  ||           :::||:
Db        552 NVKVGGINCSLIPKSNPLGNVPTLILGGDVYHPGVGATGVSIA----------SIVASV 600

Qy        634 DLQQNSTYFSTVTECS-AFDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSL 692
              || |   :: |:        :   :  ::   |: ::   :: | ||:::||| | |
Db        601 DL--NGCKYTAVSRSQPRHQEVIEGMKDIVVYLLQGFRAMTKQQPQRIIYFRDGTSEGQF 658

Qy        693 KQLFEFEVKDIIEKLKTEYARVQLSP---PQLAYIVVTRSMNTRFFLNGQ-------NPP 742
              :   |:  | |       |   |||   |::      :  : |||:  :       ||
Db        659 LSVINDELSQIKE------ACHSLSPKYNPKILVCTTQKRHHARFFIKNKSDGDRNGNPL 712

Qy        743 PGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYY 802
              ||||::  :| | :|||||:|    ||   |  | ||:  :   |:: | | | :|::|
Db        713 PGTIIEKHVTHPYQYDFYLISHPSLQGVSVPVHYTVLHDEIQMPPDQFQTLCYNLCYVYA 772

Qy        803 N-WSGTTRVPAVCQYAKKLA-----------TLVGTNLHSIPQ---------NALEKKF 840
                |  : || |  ||  ::            | | |:  |: |          : |: |
Db        773 RATSAVSLVPPV-YYAHLVSNLARYQDVTADDTFVETSEASMDQEVKPLLALSSKLKTKM 831

Qy        841 YYL 843
                :|:
Db        832 WYM 834


RESULT 5
PINH_ARATH
ID   PINH_ARATH     STANDARD;      PRT;   988 AA.
AC   Q9XGW1; O49256;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   PINHEAD protein (ZWILLE protein).
GN   AT5G43810 OR MQD19.17.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
```

```
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Landsberg erecta;
RX   MEDLINE=99094873; PubMed=9876176;
RA   Lynn K., Fernandez A., Aida M., Sedbrook J., Tasaka M., Masson P.,
RA   Barton M.K.;
RT   "The PINHEAD/ZWILLE gene acts pleiotropically in Arabidopsis
RT   development and has overlapping functions with the ARGONAUTE1 gene.";
RL   Development 126:469-481(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Landsberg erecta;
RX   MEDLINE=98169396; PubMed=9501101;
RA   Moussian B., Schoof H., Haecker A., Juergens G., Laux T.;
RT   "Role of the ZWILLE gene in the regulation of central shoot meristem
RT   cell fate during Arabidopsis embryogenesis.";
RL   EMBO J. 17:1799-1809(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: DEVELOPMENTAL PROTEIN REQUIRED FOR RELIABLE FORMATION OF
CC       PRIMARY AND AXILLARY SHOOT APICAL MERISTEMS. MAY BE A COMPONENT OF
CC       A HYPOTHETICAL MERISTEM FORMING COMPETENCE FACTOR.
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF154272; AAD40098.1; -.
DR   EMBL; AJ223508; CAA11429.1; -.
DR   EMBL; AB026651; BAB11310.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Developmental protein.
FT   DOMAIN      337    451       PAZ.
FT   DOMAIN      625    946       PIWI.
FT   CONFLICT    475    479       PRDRE -> AEGQR (IN REF. 2).
FT   CONFLICT    671    671       N -> D (IN REF. 2).
SQ   SEQUENCE   988 AA;  110867 MW;  32EBB349C613DA20 CRC64;

  Query Match             9.9%;  Score 436.5;  DB 1;  Length 988;
  Best Local Similarity   23.3%;  Pred. No. 5.8e-23;
  Matches  197;  Conservative  139;  Mismatches  346;  Indels  163;  Gaps   37;

Qy         91 GTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMGVLSNHANLLGSGYL--- 147
              || |   ::: | |     |   :  | |   | : :  |   :::    |     |
Db        135 GTLGTKCIVKANHFLADL-PTKDLNQYDVTITPEVSSKSVNRAIIAELVRLYKESDLGRR 193

Qy        148 ---FDGLQL--------FTTRKFEQEIT-----VLSGKSKLDIEYKISIKFVG------- 184
                 :|| :          || ::|  :|      :::|  | :  ||::||||
Db        194 LPAYDGRKSLYTAGELPFTWKEFSVKIVDEDDGIING-PKRERSYKVAIKFVARANMHHL 252

Qy        185 --FIS-----CAEPRFLQVLNLILRR-SMKGLNLELVGRNLFDPRAKIEIREFK-MELWP 235
                |::     | :   :|:|:::||  |:|      |||: | |  |   |  : :| |
Db        253 GEFLAGKRADCPQ-EAVQILDIVLRELSVK--RFCPVGRSFFSPDIKTPQRLGEGLESWC 309

Qy        236 GYETSIRQHEKDI----------------------LLGTEITHKVMRTETIYDIMR--R 270
                                                  |||  ::  | :      | :  |
Db        310 GFYQSIRPTQMGLSLNIDMASAAFIEPLPVIEFVAQLLGKDVLSKPLSDSDRVKIKKGLR 369

Qy        271 CSHNPARHQDEVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTF----SCKGRDISFVE 326
                    |:  ||              | ||:  :    | :  |    :|  :  | :|
Db        370 GVKVEVTHRANVR-------------RKYRVAGLTTQPTRELMFPVDENCTMK--SVIE 413

Qy        327 YYLTKYNIRIRDHNQPLLISKNRDKALKTNASELVVLIPELCR-VTGLNAEMRSNFQLMR 385
              |:   |   |:  : | |   |: ||          |  | |: | |    | | : :
Db        414 YFQEMYGFTIQHTHLPCLQVGNQKKA--------SYLPMEACKIVEGQRYTKRLNEKQIT 465

Qy        386 AMSSYTRMNPKQR-TDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIV 444
              |:   |   |: |  | ||   |   |  :     ::: | : : :  |: ||:    :
Db        466 ALLKVTCQRPRDRENDILRTVQH---NAYDQDPYAKEFGMNISEKLASVEARILPAPWLK 522

Qy        445 FH-NGK----VPAGENADWQRHFRDQRMLTTPSDGLDRWAVIAPQRNSHE--LRTLLDSL 497
              :| |||    :|  :   |  :  :::|:   :  : ||| :   |:  |   |   : |
Db        523 YHENGKEKDCLP--QVGQW--NMMNKKMINGMT--VSRWACVNFSRSVQENVARGFCNEL 576

Qy        498 YRAASGMGLRIRSPQEFIIYDDRTG--------TYVRAMDDCVRSDPKLILCLVPNDNAE 549
               :    |:          ||  |          |  :|:     : :|:| ::|::|
Db        577 GQMCEVSGMEFNPEPVIPIYSARPDQVEKALKHVYHTSMNKTKGKELELLLAILPDNNGS 636

Qy        550 RYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLGYTPWMI------EL 603
               |  :|:    :  : :|    |      |   :|  :::::| |:|    ::      :
Db        637 LYGDLKRICETELGLISQCCLTKHVFKISKQYLA-NVSLKINVKMGGRNTVLVDAISCRI 695

Qy        604 PL-SGLMTI--GFDIA--KSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTL 658
              || | : ||  | |:   ::  :   :  |::|| |  : : |   |   :    |   |
Db        696 PLVSDIPTIIFGADVTHPENGEESSPSIAAVVASQDWPEVTKYAGLVCAQAHRQELIQDL 755

Qy        659 WP-------------MIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIE 705
              :              ||   |  ::    : | ||:|||||||| |   |:  :|: | |
Db        756 YKTWQDPVRGTVSGGMIRDLLISFRKATGQKPLRIIFYRDGVSEGQFYQVLLYEL-DAIR 814

Qy        706 KLKTEYARVQLSP---PQLAYIVVTRSMNTRFFLNGQ----------NPPPGTIVDDVIT 752
              |      |   | |   | : :|||  :  :|| | |           |  |||:||  |
Db        815 K-----ACASLEPNYQPPVTFIVVQKRHHTRLFANNHRDKNSTDRSGNILPGTVVDTKIC 869

Qy        753 LPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPA 812
               |  :|||| |    |||  |  |:||:     : : :| ||  :|: |   : :  :
Db        870 HPTEFDFYLCSHAGIQGTSRPAHYHVLWDENNFTADGIQSLTNNLCYTYARCTRSVSIVP 929

Qy        813 VCQYA 817
                ||
Db        930 PAYYA 934


RESULT 6
AGOL_ARATH
ID   AGOL_ARATH     STANDARD;      PRT;   997 AA.
AC   Q9SJK3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Argonaute-like protein At2g27880.
GN   AT2G27880 OR T1E2.20.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA   Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA   Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA   Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA   Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
```

```
RA   Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AC006929; AAD21514.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
FT   DOMAIN      359    471       PAZ.
FT   DOMAIN      638    958       PIWI.
SQ   SEQUENCE   997 AA;  111134 MW;  DA93A1446C422F31 CRC64;

  Query Match             9.9%;  Score 433;  DB 1;  Length 997;
  Best Local Similarity   21.7%;  Pred. No. 1e-22;
  Matches  228;  Conservative  161;  Mismatches  377;  Indels  286;  Gaps   47;

Qy          5 QGRGRRRPXNEDDS-----------------STSRGSGDGPRVKVFRGSSSGDPRA--- 43
              :| |||    :| |                  | | | | |    | || ::||  |
Db         16 RGGGRRSDQRQDQSSGQVAWPGLQQSYGGRGGSVSAGRGRG---NVGRGENTGDLTATQV 72

Qy         44 ----------------DPRI---------------EASRERRALEEAPRREGGPPERKP 71
                            ||                 | |:     |     :    | ||
Db         73 PVASAVSGGRGRGNIGDPTFSVASSSKTVSVASSSKEESKNTEVSETMSNLQITSTETKP 132

Qy         72 WGDQYDYLNTRPVELVSK--KGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPR 129
                  ::: |    :  :|| |  ||:: | |  :   :  : || |   | : :
Db        133 EMTSLPPASSKAVTFPVRPGRGTLGKKVMVRANHFLVQV-ADRDLYHYDVSINPEVISKT 191

Qy        130 VRMGVLS------NHANLLGSGYLFDGLQ-LFTT--RKFEQEITVLS-------GKSKLD 173
              |   |:          ::| |    :|| : |:|      |: :  |::       | |  |
Db        192 VNRNVMKLLVKNYKDSHLGGKSPAYDGRKSLYTAGPLPFDSKEFVVNLAEKRADGSSGKD 251

Qy        174 IEYKISIKFVGFISCAE-PRFL------------QVLNLILRRSMKGLNLELVGRNLFDP 220
              :|:::| |     :  :||              |||:::||      :   |||: |
Db        252 RPFKVAVKNVTSTDLYQLQQFLDRKQREAPYDTIQVLDVVLRDKPSN-DYVSVGRSFFHT 310

Qy        221 RAKIEIREFKMEL------WPGYETSIRQHEKDILLGTEITHK-----VMRTETIYDIMR 269
                : |: : ||       | ||  |:|  :  : |  ::: :     :: |: |   :
Db        311 SLGKDARDGRGELGDGIEYWRGYFQSLRLTQMGLSLNIDVSARSFYEPIVVTDFISKFLN 370

Qy        270 -RCSHNPARHQDEVRV-NVLDLIVLTDYNNRTYRINDVDFGQTPKSTF----SCKGRDIS 323
               |  : | |  | ::|  ||           || ::  : :  |  :       |   |::
Db        371 IRDLNRPLRDSDRLKVKKVL----------RTLKVKLLHWNCTKSAKISGISSLPIRELR 420

Qy        324 F----------VEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASELVVLIP-ELCRVTG 372
              |          |:|:  ||| |::    |          |::| :    | :| |||::
Db        421 FTLEDKSEKTVVQYFAEKYNYRVKYQALP---------AIQTGSDTRPVYLPMELCQID- 470

Qy        373 LNAEMRSNFQLMRAMSSYT-RMNPKQRTDRLRAFNHRLQNTPESVKVL----------R 420
                           || |:| || |  |:|   |  :   |:| |           :
Db        471 -------------EGQRYTKRLNEKQVTALLKATCQRPPDRENSIKNLVVKNNYNDDLSK 517

Qy        421 DWNMELDKNVTEVQGRIIGQQNIVFH-NGKVPAGENADWQRHFRDQRMLTTPSDGLDRWA 479
              :: | :   :  :: |::    : :| :||         | :  |::|:      :  |
Db        518 EFGMSVTTQLASIEARVLPPPMLKYHDSGKEKMVNPRLGQWNMIDKKMVNGAK--VTSWT 575

Qy        480 VIA---------PQRNSHELRTLLDSLYRAASGMGLR-------IRSPQEFIIYDDRTGT 523
              ::         ||    :|  :       : ||  :       |  | | |
```

```
Db        576 CVSFSTRIDRGLPQEFCKQLIGMC-----VSKGMEFKPQPAIPFISCPPEHI-------- 622

Qy        524 YVRAMDDCVRSDP--KLILCLVPNDNAERYSSIKK-----RGYVDRAV-PTQVVTLKTTK 575
                |: |  :  |  :|:: ::| |    |  ||:      | | :   | ||  |
Db        623 -EEALLDIHKRAPGLQLLIVILP-DVTGSYGKIKRICETELGIVSQCCQPRQVNKL---- 676

Qy        576 NRSLMSIATKIAIQLNCKLG-----------YTPWMIELPLSGLMTIGFDIA--KSTRD 621
              |: |     :|:::| | |                | : : |    : :| |:   :  |
Db        677 NKQYME---NVALKINVKTGGRNTVLNDAIRRNIPLITDRP---TIIMGADVTHPQPGED 730

Qy        622 RKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQH-----EH--- 673
                :  |::|||| : : |   |:  :  : :   |: ::    |  |     ||
Db        731 SSPSIAAVVASMDWPEINKYRGLVSAQAHREEIIQDLYKLVQDPQRGLVHSGLIREHFIA 790

Qy        674 -----RKLPSRIVFYRDGVSSGSLKQLFEFE---VKDIIEKLKTEYARVQLSPPQLAYIV 725
                   ::| ||:|||||| |'  |:   |   ::     |:  |       |:: :::
Db        791 FRRATGQIPQRIIFYRDGVSEGQFSQVLLHEMTAIRKACNSLQENYV------PRVTFVI 844

Qy        726 VTRSMNTRFFL----------NGQNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTS 775
              | :  :|| |            |  |||:||  |  |  :|||| |    |||  |
Db        845 VQKRHHTRLFPEQHGNRDMTDKSGNIQPGTVVDTKICHPNEFDFYLNSHAGIQGTSRPAH 904

Qy        776 YNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYA----------------- 817
                |:||    | : :::| ||  :|: |   : :  :     ||
Db        905 YHVLLDENGFTADQLQMLTNNLCYTYARCTKSVSIVPPAYYAHLAAFRARYYMESEMSDG 964

Qy        818 -----KKLATLVG---TNLHSIPQNALEKKFY 841
                    :   | ||   : | :|  |  |  ||
Db        965 GSSRSRSSTTGVGQVISQLPAIKDNVKEVMFY 996


RESULT 7
AGO1_ARATH
ID   AGO1_ARATH     STANDARD;      PRT;  1048 AA.
AC   O04379;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Argonaute protein.
GN   AGO1 OR AT1G48410 OR F11A17.3.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia; TISSUE=Leaf;
RX   MEDLINE=98090460; PubMed=9427751;
RA   Bohmert K., Camus I., Bellini C., Bouchez D., Caboche M., Benning C.;
RT   "AGO1 defines a novel locus of Arabidopsis controlling leaf
RT   development.";
RL   EMBO J. 17:170-180(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
```

```
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).
CC   -!- FUNCTION: ESSENTIAL FOR PROPER DEVELOPMENT OF LEAVES AND FLORAL
CC       ORGANS, AND FORMATION OF AXILLARY MERISTEMS.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED AT LOW LEVELS.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT ALL DEVELOPMENTAL
CC       STAGES.
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U91995; AAC18440.1; -.
DR   EMBL; AC007932; AAD49755.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Developmental protein.
FT   DOMAIN      391    501       PAZ.
FT   DOMAIN      676    997       PIWI.
FT   DOMAIN       13    104       GLY-RICH.
SQ   SEQUENCE   1048 AA;  116190 MW;  3E5146343A09C541 CRC64;

  Query Match             9.8%;  Score 431;  DB 1;  Length 1048;
  Best Local Similarity   22.7%;  Pred. No. 1.6e-22;
  Matches  217;  Conservative  149;  Mismatches  396;  Indels  194;  Gaps   42;

Qy          5 QGRGRRRPXNEDD-SSTSRGSGDGPRVKVFRGSSSGDPR----------ADPRIEASRER 53
              |||||  | ::       | | ||       ||| |:            |  :|   :
Db         81 QGRGRGGPPHQGGRGGYGGGRGGGP--------SSGPPQRQSVPELHQATSPTYQAVSSQ 132

Qy         54 RALEEAPRREGGPPERKPWGDQYDYLNT---------RPVELVSK---------KGTDGV 95
               | |    :   ||      |:: |:           :|:   ||         ||  |
Db        133 PTLSEVSPTQ--VPEPTVLAQQFEQLSVEQGAPSQAIQPIPSSSKAFKFPMRPGKGQSGK 190

Qy         96 PVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMGVLSNHANLLGSGYL------FD 149
                 ::: | |  :  |:  : || |   | : :  |   |:    :     :|      :|
Db        191 RCIVKANHFFAEL-PDKDLHHYDVTITPEVTSRGVNRAVMKQLVDNYRDSHLGSRLPAYD 249

Qy        150 GLQL--------FTTRKFEQEI---TVLSGKSKLDIEYKISIKFV-------------GF 185
               | :         | :::|   :    | :|  : : |:|: || |             |
Db        250 GRKSLYTAGPLPFNSKEFRINLLDEEVGAGGQRREREFKVVIKLVARADLHHLGMFLEGK 309

Qy        186 ISCAEPRFLQVLNLILRRSMKGLNLEL-------VGRNLFDPR-AKIEIREFKMELWPGY 237
               | |    ||||:::||        ||       |||: : |   | :     :| | |:
Db        310 QSDAPQEALQVLDIVLR--------ELPTSRYIPVGRSFYSPDIGKKQSLGDGLESWRGF 361

Qy        238 ETSIRQHEKDILLGTEITHKVM-----RTETIYDIMRR-CSHNPARHQDEVRVNVL---D 288
               |||  :  : |  :::          : : |:: |  |  |    | |::
Db        362 YQSIRPTQMGLSLNIDMSSTAFIEANPVIQFVCDLLNRDISSRPLSDADRVKIKKALRGV 421

Qy        289 LIVLTDYNN--RTYRINDVDFGQTPKSTFSCKGRDI--SFVEYYLTKYNIRIRDHNQPLL 344
               : :|   |  | |||: :    | : ||    |:   | |||:   |  ||:    | |
Db        422 KVEVTHRGNMRRKYRISGLTAVATRELTFPVDERNTQKSVVEYFHETYGFRIQHTQLPCL 481

Qy        345 I--SKNRDKALKTNASELVVLIPELCR-VTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDR 401
               : ||  |            |:|: | |     | | : : |:   |   |    ||
Db        482 QVGNSNRPNYLPM----------EVCKIVEGQRYSKRLNERQITALLKVTCQRP---IDR 528

Qy        402 LRAFNHRLQ-NTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFH-NGKVPAGENADWQ 459
               :     :| |        ::: :::  ::  |: ||:    : :| :|:          |
Db        529 EKDILQTVQLNDYAKDNYAQEFGIKISTSLASVEARILPPPWLKYHESGREGTCLPQVGQ 588

Qy        460 RHFRDQRMLTTPSDGLDRWAVI--APQRNSHELRTLLDSLYRA--ASGMGLRIRSPQEFI 515
               :  :::|:      :: |  |  : |   :  || .  | :    |||     :|:  :
Db        589 WNMMNKKMIN--GGTVNNWICINFSRQVQDNLARTFCQELAQMCYVSGMAF---NPEPVL 643

Qy        516 ---------------IYDDRTGTYVRAMDDCVRSDPKLILCLVPNDNAERYSSIKKRGY 559
                            | | |    :        :  |:: ::|::|   |  :|:
Db        644 PPVSARPEQVEKVLKTRYHDATSKLSQG------KEIDLLIVILPDNNGSLYGDLKRICE 697

Qy        560 VDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLGYTPWMI------ELPL-SGLMTI- 611
               :  : :|    |     |   :|  :|:::| |:|    ::       :|| |   ||
Db        698 TELGIVSQCCLTKHVFKMSKQYMA-NVALKINVKVGGRNTVLVDALSRRIPLVSDRPTII 756

Qy        612 -GFDIA--KSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTL---WP----- 660
               | |:       |   :  |::|| |  : : |   |    :    |   |   |
Db        757 FGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFKEWKDPQKG 816

Qy        661 -----MIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEK----LKTEY 711
                  || : |  ::      | ||:|||||| |   |:  :|: | | |    |:  |
Db        817 VVTGGMIKELLIAFRRSTGHKPLRIIFYRDGVSEGQFYQVLLYEL-DAIRKACASLEAGY 875

Qy        712 ARVQLSPPQLAYIVVTRSMNTRFFLNGQNPP----------PGTIVDDVITLPERYDFYL 761
                  | : ::|| :  :|| |    |             |||:||  |  |  :||||
Db        876 ------QPPVTFVVVQKRHHTRLFAQNHNDRHSVDRSGNILPGTVVDSKICHPTEFDFYL 929

Qy        762 VSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYA 817
               |    |||  |  |:||:      : : :| ||  :|: |   : :  :     ||
Db        930 CSHAGIQGTSRPAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSIVPPAYYA 985


RESULT 8
YO43_CAEEL
ID   YO43_CAEEL     STANDARD;      PRT;  1040 AA.
AC   P34681;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical 115.4 kDa protein ZK757.3 in chromosome III.
GN   ZK757.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA   Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA   Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA   Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA   Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA   Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA   Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; Z30215; CAA82941.1; -.
DR   EMBL; Z29121; CAA82941.1; JOINED.
DR   EMBL; Z29121; CAA82389.1; -.
DR   EMBL; Z30215; CAA82389.1; JOINED.
DR   PIR; S41013; S41013.
DR   WormPep; ZK757.3; CE01117.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Hypothetical protein.
FT   DOMAIN      378    486       PAZ.
FT   DOMAIN      660    966       PIWI.
SQ   SEQUENCE   1040 AA;  115415 MW;  4C8483C1F1D72338 CRC64;

  Query Match             9.8%;  Score 429.5;  DB 1;  Length 1040;
  Best Local Similarity   20.4%;  Pred. No. 2e-22;
  Matches  212;  Conservative  159;  Mismatches  363;  Indels  303;  Gaps   43;

Qy          9 RRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG--- 65
              ||    :  |::|   ||          || |  |   |  :       : :|    |
Db          3 RRNATSFVDNNTLTSSG-------ISGSGSMSPPITSRPASG------QASPLTSNGSLS 49

Qy         66 PPERKPWGDQYDY-----------------LNTRPVELVSKKGTDGVPVMLQTNFFRXK 107
               ||:          |                || | |      || |  : :::|||
Db         50 PPQYADDQGSVSYNLDSPRDLSPLLLSELACLNMREVVARPGLGTIGRQIPVKSNFFAMD 109

Qy        108 TK-PEWRIVHYHVEFVPTIENPRVR----------MGVLSNHANLLGSGY--LFDGL-Q 152
               | |:  :: ||||    | :|  |             :|:| |:  : :   :||  |
Db        110 LKNPKMVVIQYHVE----IHHPGCRKLDKDEMRIIFWKAVSDHPNIFHNKFALAYDGAHQ 165

Qy        153 LFTTRKFEQEITVLSGKSKLDIEYK-----------ISIKFVGFISCAEPRF------- 193
              |:|  :   |      |  :|| |             |||: || :     |
Db        166 LYTVARL--EFPDDQGSVRLDCEASLPKDNRDRTRCAISIQNVGPVLLEMQRTRTNNLDE 223

Qy        194 -----LQVLNLILRRSM----------------------KGLNLELVGRNLFDPRAKI 224
                  :|:|::| |:|:                      |  |:| |
Db        224 RVLTPIQILDIICRQSLTCPLLKNSANFYTWKSSCYRIPTAAGQALDLEGGK-------- 275

Qy        225 EIREFKMELWPGYETS--IRQHEKDILLGTEITH------KVMRTETIYDIMRRCSHNPA 276
                  |:| |: :|  |  : :  ||  :: |       ::   : : |::   :  |
Db        276 -------EMWTGFFSSAHIASNYRP-LLNIDVAHTAFYKTRITVLQFMCDVLNERTSKPN 327

Qy        277 RH--------------------------------------QDEVRVNVLDLIV------ 291
              |:                                     :|:   | |   :
Db        328 RNNPRGPGGPGGPGGYRGGRGGGRGGSYGNFGNRGPPGANVRDDFGGNGLTFTMDTLSRD 387

Qy        292 --LTDYNNRTY-----------------RINDVDFGQTPKSTFSCKGRD-------ISF 324
                |: :  | :                |:  |:  | |       :| |          |
Db        388 TQLSSFETRIFGDAIRGMKIRAAHRPNAIRVYKVNSLQLPADKLMFQGIDEEGRQVVCSV 447

Qy        325 VEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASELVVLIP-ELCRVTGLNAEMRSNFQL 383
               :|:  ||          ||   |      |          : :| | | :   ::  : | ::
Db        448 ADYFSEKYG--------PLKYPKL--PCLHVGPPTRNIFLPMEHCLI---DSPQKYNKKM 494

Qy        384 MRAMSS----YTRMNPKQRTDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIG 439
               :|       ::  || ||::      | :  :   |::: : :   : :   |:|
Db        495 SEKQTSAIIKAAAVDATQREDRIKQL--AAQASFGTDPFLKEFGVAVSSQMIQTTARVIQ 552

Qy        440 QQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDG---LDRWAVIAP------------ 483
              |:|       | |       |    :  | ||   :|   :  |
Db        553 PPPIMF------GGNN-------RSVNPVVFPKDGSWTMDNQTLYMPATCRSYSMIALVD 599

Qy        484 QRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIY----DDRTGTYVRAMD---------D 530
               |:   |:|   ||   |: ||:       : : |     :|     :     |
Db        600 PRDQTSLQTFCQSLTMKATAMGMNFPRWPDLVKYGRSKEDVCTLFTEIADEYRVTNTVCD 659

Qy        531 CVRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQL 590
              |       |: :: : |:: | ::|::  :    : :| | :|   :|   :     | ::|
Db        660 C-------IIVVLQSKNSDIYMTVKEQSDIVHGIMSQCVLMKNV-SRPTPATCANIVLKL 711

Qy        591 NCKLG-----------YTPWMIELPLSGLMTIGFDIAKSTRDRKR----AYGALIASMDL 635
              | |:|              :::: |    | :| |:   |:   |    :  |::|::||
Db        712 NMKMGGINSRIVADKITNKYLVDQP---TMVVGIDVTHPTQAEMRMNMPSVAAIVANVDL 768

Qy        636 QQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQL 695
                  |   :   :   : :   |  :  : :  |:|  :| |:||: |||||  |    ::
Db        769 LPQSYGANVKVQKKCRESVVYLLDAIRERIITFYRHTKQK-PARIIVYRDGVSEGQFSEV 827

Qy        696 FEFEVKDIIEKLKTEYARV--QLSPPQLAYIVVTRSMNTRFFLNGQNP--------PPGT 745
                ::: |: ::|    :     || : |||| :  : | |   ||            ||||
Db        828 ----LREEIQSIRTACLAIAEDFRPP-ITYIVVQKRHHARIFCKYQNDMVGKAKNVPPGT 882

Qy        746 IVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWS 805
              ||  |  || :|||| |    |||  |  |:||       : :::| :|| ||| |   :
Db        883 TVDTGIVSPEGFDFYLCSHYGVQGTSRPARYHVLLDECKFTADEIQSITYGMCHTYGRCT 942

Qy        806 GTTRVPAVCQYAKKLAT 822
               :  :|    ||  :||
Db        943 RSVSIPTPVYYADLVAT 959


RESULT 9
I2C2_HUMAN
ID   I2C2_HUMAN     STANDARD;      PRT;   377 AA.
AC   Q9UKV8;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2)
DE   (Fragment).
GN   EIF2C2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20005943; PubMed=10534406;
RA   Koesters R., Adams V., Betts D., Moos R., Schmid M., Siermann A.,
RA   Hassam S., Weitz S., Lichter P., Heitz P.U., von Knebel Doeberitz M.,
RA   Briner J.;
RT   "Human eukaryotic initiation factor EIF2C1 gene: cDNA sequence,
RT   genomic organization, localization to chromosomal bands 1p34-p35, and
RT   expression.";
RL   Genomics 61:210-218(1999).
CC   -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC       initiation process.
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF121255; AAF13034.2; -.
DR   Genew; HGNC:3263; EIF2C2.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Initiation factor; Protein biosynthesis.
FT   NON_TER       1      1
FT   DOMAIN       35    336       PIWI.
SQ   SEQUENCE   377 AA;  42529 MW;  3BBCE91689D58526 CRC64;

  Query Match             7.3%;  Score 321;  DB 1;  Length 377;
  Best Local Similarity   29.3%;  Pred. No. 2e-15;
  Matches   90;  Conservative   58;  Mismatches  125;  Indels   34;  Gaps    9;

Qy        537 KLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLGY 596
```

```
                 :|:: ::|      |: :|: |     : || | :|  : |:     : : :::| |||
Db         34 QLVVVILPG-KTPVYAEVKRVGDTVLGMATQCVQMKNVQ-RTTPQTLSNLWLKINVKLGG 91

Qy        597 T---------PWMIELPLSGLMTIGFDIA--KSTRDRKRAYGALIASMDLQQNSTYFSTV 645
                        | : : |   :: :| |:    :   :| :  |::  ||   |  | :||
Db         92 VNNILLPQGRPPVFQEP---VIFLGADVTHPPAGDGKKPSIAAVVGRMDAHPN-RYCATV 147

Qy        646 TECSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIE 705
                      :   |  |: : | |:    |  |:||:|||||| |  :|:   |:  | |
Db        148 RVQQHRQEIIQDLAAMVRELLIQFYKSTRFKPTRIIFYRDGVSEGQFQQVLHHELLAIRE 207

Qy        706 ---KLKTEYARVQLSPPQLAYIVVTRSMNTRFFLNGQNP--------PPGTIVDDVITLP 754
                 ||: :|       | : :||| :  :|| |   :|          | || ||  || |
Db        208 ACIKLEKDY------QPGITFIVVQKRHHTRLFCTDKNERVGKSGNIPAGTTVDTKITHP 261

Qy        755 ERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVC 814
                :|||| |    |||  |: |:||:     | :::| |||::|| |   : :  :||
Db        262 TEFDFYLCSHAGIQGTSRPSHYHVLWDDNRFSSDELQILTYQLCHTYVRCTRSVSIPAPA 321

Qy        815 QYAKKLA 821
                ||  :|
Db        322 YYAHLVA 328
```

RESULT 10
YQ53_CAEEL
ID YQ53_CAEEL STANDARD; PRT; 1032 AA.
AC Q09249;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C16C10.3 in chromosome III.
GN C16C10.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Lloyd C.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: STRONG, TO C.ELEGANS C14B1.7. ALSO TO C06A1.4,
CC F48F7.1, R06C7.1, T22B3.2 AND ZK757.3.
CC -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; Z46787; CAA86741.1; -.
DR WormPep; C16C10.3; CE01494.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
KW Hypothetical protein.
FT DOMAIN 369 481 PAZ.
FT DOMAIN 650 977 PIWI.
SQ SEQUENCE 1032 AA; 117105 MW; EBEAE23CE67BB149 CRC64;

 Query Match 5.8%; Score 256.5; DB 1; Length 1032;
 Best Local Similarity 22.2%; Pred. No. 3.5e-10;
 Matches 145; Conservative 98; Mismatches 259; Indels 151; Gaps 28;

```
Qy        269 RRCSHNPARHQDEVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTFSCKG----RDISF 324
                 :||      :  |         :  |       |:: :       ::|  |     |:::
Db        391 KRCIEEALKGLD---------VECTHLKGNLIRVSSIAENNAENTSFMMKDDKGEREVTV 441

Qy        325 VEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASELVVLIP-----------ELCRVTG 372
                ||:| :|||:::    ||:::|| | |       ||: : |                :||
Db        442 AEYFLLQYNIKLKYPRLPLVVSK-RFKHESFFPMELLRIAPGQRIKVNKMSPTVQSAMTG 500

Qy        373 LNAEMRSNFQLMRAMSSYTRMNPK-QRTDRLRAFNHRLQNT----------PESVK--- 417
                || |  :   :: :    | | | ::   : ||  :| :|          |  :|
Db        501 RNASMPQHH--VKLVQDILRDNLKLEQNKYMDAFGIKLMSTEPIQMTAKLLPPAQIKFKG 558

Qy        418 ------VLRDWNMELDKNVTEVQGRIIGQQNIVFHN--GKVPAGENADWQRHFRDQRMLT 469
                     : |      || |   : | ||   :|| |      | : |   :|     :|
Db        559 QTYMPDMSRPAFRTQDKFVEPARIRKIG--IVVFDNCIQMRQAEDFCDKLSNFCRDNGIT 616

Qy        470 TPSDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMD 529
                  | | |::   : || :    : :| :                          : :|
Db        617 VEKDSRD-WSI--RELNSSD-SVAIQNLMK-------------------------KWLD 646

Qy        530 DCVRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTL------KTTKNRSLMSIA 583
                |  | |   ||  :  :       | |  | : ::  | : |      |    :
Db        647 D--RVD---ILVGIAREKKPDVHDILK--YFEESIGLQTIQLCQQTVDKMMGGQGGRQTI 699

Qy        584 TKIAIQLNCKLGYTPWMIELP------------------LSGLMTIGFDI----AKSTR 620
                  :. : | | | | : :|:|                |  :  |||:|    :::
Db        700 DNVMRKFNLKCGGTNFFVEIPNAVRGKAVCSNNETLRKKLLEHVQFIGFEISHGASRTLF 759

Qy        621 DRKRAYGALIASMDLQQN----STYFSTVTECSAFDVL-------ANTLWPMIAKALRQY 669
                ||  |      : || : :  |  :  |:   |  |          |     | :| |
Db        760 DRSR------SQMDGEPSVVGVSYSLTNSTQLGGFTYLQTQKEYKLQKLDEFFPKCVRSY 813

Qy        670 QHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLS-PPQLAYIVVTR 728
                :   : ||:|:|| || |   |:  :     ||: :|:::  : ::|    | |  |:  |
Db        814 KEHSKTLPTRIVIYRVGAGEGNFNR-----VKEEVEEMRRTFDKIQPGYRPHLVVIAQR 868

Qy        729 SMNTRFF---LNG-----QNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLY 780
                : : | |   ::|     || | || |::|:|     :| | ||    ||| |  | :|
Db        869 ASHARVFPSCISGNRATDQNIPSGTCVENVLTSYGYDEFILSSQTPLIGTVRPCKYTILV 928

Qy        781 SSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQ 833
                :   |  ::  |||     :        || |   |: ||   | | : | |
Db        929 NDAKWSKNELMHLTYFRAFGHQVSYQPPSVPDVLYAAENLAKR-GRNNYKIHQ 980
```

RESULT 11
RA50_PYRFU
ID RA50_PYRFU STANDARD; PRT; 882 AA.
AC P58301;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR PF1167.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RX MEDLINE=20485553; PubMed=11029422;
RA Hopfner K.-P., Karcher A., Shin D., Fairley C., Tainer J.A.,
RA Carney J.P.;
RT "Mre11 and Rad50 from Pyrococcus furiosus: cloning and biochemical
RT characterization reveal an evolutionarily conserved multiprotein
RT machine.";
RL J. Bacteriol. 182:6036-6041(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.

```
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 1-149.
RC   STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RX   MEDLINE=20348838; PubMed=10892749;
RA   Hopfner K.-P., Karcher A., Shin D.S., Craig L., Arthur L.M.,
RA   Carney J.P., Tainer J.A.;
RT   "Structural biology of Rad50 ATPase: ATP-driven conformational
RT   control in DNA double-strand break repair and the ABC-ATPase
RT   superfamily.";
RL   Cell 101:789-800(2000).
CC   -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC       rad50/mre11 complex possesses single-strand endonuclease activity
CC       and ATP-dependent double-strand-specific exonuclease activity.
CC       Rad50 provides an ATP-dependent control of mre11 by unwinding
CC       and/or repositioning DNA ends into the mre11 active site.
CC   -!- SUBUNIT: Forms a complex with mre11.
CC   -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE010225; AAL81291.1; -.
DR   PDB; 1F2T; 02-AUG-00.
DR   PDB; 1F2U; 02-AUG-00.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR003395; SMC_N.
DR   Pfam; PF02463; SMC_N; 1.
KW   DNA repair; Hydrolase; ATP-binding; Coiled coil; 3D-structure;
KW   Complete proteome.
FT   NP_BIND      30     37       ATP.
FT   DOMAIN      148    744       COILED COIL (POTENTIAL).
SQ   SEQUENCE   882 AA;  103839 MW;  3ADCBD250382A99E CRC64;

  Query Match             3.0%;  Score 129.5;  DB 1;  Length 882;
  Best Local Similarity   21.4%;  Pred. No. 0.27;
  Matches   67;  Conservative   68;  Mismatches  113;  Indels   65;  Gaps   15;

Qy        161 QEITVLSGKSKLD--IEYKISIKFVGFISCAEPRFLQVLNLILRRSMKG----LNLELVG 214
              : :|| : :|  |  :|:|                 : :|||: ::  |    |:  |||
Db          4 ERVTVKNFRSHSDTVVEFK----------------EGINLIIGQNGSGKSSLLDAILVG 46

Qy        215 RNLFDPRAKIEIREFKMELWPGYETSIRQHEKDILL---GTE--ITHKVMRTETIYDI-- 267
              |: |   : |:: | : :   :   |    |::    ||:  || : ::  :  :|
Db         47 --LYWP---LRIKDIKKDEFT--KVGARDTYIDLIFEKDGTKYRITRRFLKGYSSGEIHA 99

Qy        268 MRRCSHNPARHQDEVRVNVLDLIV--LTDYN---NRTY-RINDVDFGQTPKSTFSCKGRD 321
              |:|   |  :|  |     :   :  |  ||   |  | |   :|              |:
Db        100 MKRLVGNEWKHVTEPSSKAISAFMEKLIPYNIFLNAIYIRQGQIDAILESDEAREKVVRE 159

Qy        322 ISFVEYYLTKY----------NIRIRDHNQPLLISKNRDKALKTNASELVVLIPELCRVT 371
              :  :: : | |          | ||:::   |  ::| :: :| |  ||: :: |: ::
Db        160 VLNLDKFETAYKKLSELKKTINNRIKEYRDILARTENIEELIKENEQELIQVLQEISKIE 219

Qy        372 GLNAEMRSNFQLMR----------AMSSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLR 420
              :    ||   ::|          : :  |:  |:| |: |    |::|| | :: |:
Db        220 EVLPSKRSKVDMLRKEVLRLEETKVEIENSERLLEKRRGDK-RTLEERIKNTEEYLEKLK 278

Qy        421 DWNMELDKNVTEV 433
              :   ||:: | |:
Db        279 EKEKELEEQVKEI 291


RESULT 12
FAT_DROME
ID   FAT_DROME      STANDARD;      PRT;  5147 AA.
AC   P33450; Q9VQX5;
DT   01-FEB-1994 (Rel. 28, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Cadherin-related tumor suppressor precursor (Fat protein).
GN   FT OR CG3352.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92069752; PubMed=1959133;
RA   Mahoney P.A., Weber U., Onofrechuk P., Biessmann H., Bryant P.J.,
RA   Goodman C.S.;
RT   "The fat tumor suppressor gene in Drosophila encodes a novel member
RT   of the cadherin gene superfamily.";
RL   Cell 67:853-868(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: COULD FUNCTION AS A CELL-ADHESION PROTEIN. ACTS AS A
CC       TUMOR SUPPRESSOR. REQUIRED FOR CORRECT MORPHOGENESIS.
CC   -!- SIMILARITY: CONTAINS 34 CADHERIN DOMAINS.
CC   -!- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M80537; AAA28530.1; -.
```

```
DR   EMBL; AE003577; AAF51036.1; -.
DR   PIR; A41087; IJFFTM.
DR   HSSP; P00740; 1EDM.
DR   FlyBase; FBgn0001075; ft.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00028; cadherin; 34.
DR   Pfam; PF00054; laminin_G; 2.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 34.
DR   SMART; SM00181; EGF; 4.
DR   SMART; SM00282; LamG; 2.
DR   PROSITE; PS00232; CADHERIN_1; 22.
DR   PROSITE; PS50268; CADHERIN_2; 34.
DR   PROSITE; PS00022; EGF_1; 4.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
KW   Cell adhesion; Signal; Transmembrane; Glycoprotein; Calcium-binding;
KW   Repeat; EGF-like domain.
FT   SIGNAL        1     35       POTENTIAL.
FT   CHAIN        36   5147       CADHERIN-RELATED TUMOR SUPPRESSOR.
FT   DOMAIN       36   4583       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   4584   4609       POTENTIAL.
FT   DOMAIN     4610   5147       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       36    156       CADHERIN 1.
FT   DOMAIN      157    270       CADHERIN 2.
FT   DOMAIN      271    382       CADHERIN 3.
FT   DOMAIN      383    494       CADHERIN 4.
FT   DOMAIN      495    599       CADHERIN 5.
FT   DOMAIN      600    708       CADHERIN 6.
FT   DOMAIN      709    820       CADHERIN 7.
FT   DOMAIN      821    942       CADHERIN 8.
FT   DOMAIN      943   1049       CADHERIN 9.
FT   DOMAIN     1050   1153       CADHERIN 10.
FT   DOMAIN     1154   1278       CADHERIN 11.
FT   DOMAIN     1279   1384       CADHERIN 12.
FT   DOMAIN     1385   1489       CADHERIN 13.
FT   DOMAIN     1490   1601       CADHERIN 14.
FT   DOMAIN     1602   1713       CADHERIN 15.
FT   DOMAIN     1714   1823       CADHERIN 16.
FT   DOMAIN     1824   1922       CADHERIN 17.
FT   DOMAIN     1923   2027       CADHERIN 18.
FT   DOMAIN     2028   2167       CADHERIN 19.
FT   DOMAIN     2168   2278       CADHERIN 20.
FT   DOMAIN     2279   2385       CADHERIN 21.
FT   DOMAIN     2386   2491       CADHERIN 22.
FT   DOMAIN     2492   2596       CADHERIN 23.
FT   DOMAIN     2597   2703       CADHERIN 24.
FT   DOMAIN     2704   2810       CADHERIN 25.
FT   DOMAIN     2811   2913       CADHERIN 26.
FT   DOMAIN     2914   3013       CADHERIN 27.
FT   DOMAIN     3014   3124       CADHERIN 28.
FT   DOMAIN     3125   3229       CADHERIN 29.
FT   DOMAIN     3230   3334       CADHERIN 30.
FT   DOMAIN     3335   3439       CADHERIN 31.
FT   DOMAIN     3440   3545       CADHERIN 32.
FT   DOMAIN     3546   3651       CADHERIN 33.
FT   DOMAIN     3652   3756       CADHERIN 34.
FT   DOMAIN     3950   4011       EGF-LIKE 1.
FT   DOMAIN     4013   4049       EGF-LIKE 2.
FT   DOMAIN     4052   4090       EGF-LIKE 3.
FT   DOMAIN     4092   4128       EGF-LIKE 4.
FT   DOMAIN     4129   4320       LAMININ G-LIKE 1.
FT   DOMAIN     4321   4362       EGF-LIKE 5.
FT   DOMAIN     4402   4569       LAMININ G-LIKE 2.
FT   DISULFID   3954   3966       BY SIMILARITY.
FT   DISULFID   3960   3999       BY SIMILARITY.
FT   DISULFID   4001   4010       BY SIMILARITY.
FT   DISULFID   4017   4028       BY SIMILARITY.
FT   DISULFID   4022   4037       BY SIMILARITY.
FT   DISULFID   4039   4048       BY SIMILARITY.
FT   DISULFID   4056   4067       BY SIMILARITY.
FT   DISULFID   4061   4078       BY SIMILARITY.
FT   DISULFID   4080   4089       BY SIMILARITY.
FT   DISULFID   4096   4107       BY SIMILARITY.
FT   DISULFID   4101   4116       BY SIMILARITY.
FT   DISULFID   4118   4127       BY SIMILARITY.
FT   DISULFID   4325   4341       BY SIMILARITY.
FT   DISULFID   4334   4350       BY SIMILARITY.
FT   DISULFID   4352   4361       BY SIMILARITY.
FT   CARBOHYD    239    239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    257    257       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    276    276       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    280    280       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    402    402       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    461    461       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    605    605       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    631    631       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1155   1155       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1367   1367       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1458   1458       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1751   1751       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1831   1831       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1880   1880       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2080   2080       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2171   2171       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2247   2247       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2290   2290       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2437   2437       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2581   2581       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2799   2799       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2920   2920       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2946   2946       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2967   2967       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3167   3167       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3303   3303       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3386   3386       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3389   3389       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3525   3525       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3852   3852       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3865   3865       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3905   3905       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4306   4306       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4414   4414       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4471   4471       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4487   4487       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4539   4539       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4550   4550       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT    1229   1229       S -> G.
FT   VARIANT    1233   1233       G -> S.

  Query Match              2.8%;  Score 124.5;  DB 1;  Length 5147;
  Best Local Similarity   20.4%;  Pred. No. 9.2;
  Matches  205;  Conservative  129;  Mismatches  394;  Indels  279;  Gaps   52;

Qy         29 RVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGGPPERKPWGDQYDYLNTRPVEL-- 86
              |::|  |    :  |  ::  : :   | |        ||     | |   :  | :|
Db       1894 RIRVSDGVQYTEAYATIQVSDTNDNPPLFEDTVYSFDIPENAQRGYQVGQIVARDADLGQ 1953

Qy         87 --------VSKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMG----V 134
                     ||    |   : ||       : ::  | :::  |   :| :  :     |
Db       1954 NAQLSYGVVSDWANDVFSLNPQTGMLTLTARLDYEEVQHYILIVQAQDNGQPSLSTTITV 2013

Qy        135 LSNHANLLGSGYLFDGLQLFTTRKFEQ-----EITVLSGKSKLD------IEYKISIKFV 183
              |  :|  :  :|| :  ::: ||      |:  :| |  :|      ||| |:   |
Db       2014 YCNVLDLNDNAPIFDPMS-YSSEVFENVPIATEVVTVSAKD-IDSGNNGLIEYSITAGDV 2071

Qy        184 ----GFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKI-EIREFKMEL---WP 235
                  |  |    |  : |:  |       | :  |:  |  |   |: | :::|    |
Db       2072 DSEFGIDSNGTIRTRRNLD---REHRSTYTLTVTARDCADEFASFSELEETQLKLKYRSP 2128

Qy        236 GYETSIRQ----HEKDILLGTEITHKVMRTETIYDIMRRCSHNPARHQDEVRVNVLDLIV 291
                 ||    |:|   | : :   ::  :   ::    | |     : | :| : : |
```

```
Db       2129 RKYQQTRQEFLAHQKQQRLSSTVKVTILIKDVNDEVPVFISANETAIMENVAINTVVIAV 2188

Qy        292 LTDYNNRTYRINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKN-RD 350

Db       2189 --------------------KAVDNDEGRN-GYIDYLMKEARDEDMGQSDPLPFSLNPTD 2227

Qy        351 KALKTNASELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDR---LRAFNH 407

Db       2228 GQLRV--------------VDALDRELRSSYLL--NITARDRGEPPQSTESQLLIRILDE 2271

Qy        408 RLQNTP------ESVKVLRDWN---MELDKNVTEV----QGRI-----IGQQNIVF---- 445

Db       2272 N-DNSPVFDPKQYSASVAENASIGAMVLQVSATDVDEGANGRIRYSIVLGDQNHDFSISE 2330

Qy        446 HNGKV--------------------------PAGENA----------DWQRHFRDQRML 468

Db       2331 DTGVVRVAKNLNYERLSRYSLTVRAEDCALENPAGDTAELTINILDINDNRPTFLDSPYL 2390

Qy        469 ------TTPSDG-----LDRWAVIAPQRNSHELRTLL----DSLYRA-ASGMGLRIRSPQ 512

Db       2391 ARVMENTVPPNGGYVLTVNAYDADTPPLNS-QVRYFLKEGDSDLFRINASSGDIALLKPL 2449

Qy        513 EFIIYDDRTGTYVRAMD---------DCVRSDPKLILCLVPNDNAERYSSIKKRGYVDRA 563

Db       2450 DREQQSEYTLTLV-AMDTGSPPLTGTGIVRVEVQDI-----NDNDPVFELQSYHATVREN 2503

Qy        564 VP--TQVVTLKTTKNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRD 621

Db       2504 LPSGTHVLTPRATDKDE----------GLNAKLRFN-------LLGEHMHRFHIDSETGE 2546

Qy        622 RKRAYGALIASMDLQQNSTYF-------STVTE---------CSAFDVLANTLWPMIAKA 665

Db       2547 ISTA-----TTLDREETSVYHLTLMAQDSSITEPRASSVNLTISVSDVNDNI--PKFDST 2599

Qy        666 LRQYQHEHRKLPSRIVF------YRDGVS-------SGSLKQLFEFEVKDIIEKLKTEYA 712

Db       2600 TYNVAVPERISKGEFVFGARALDLDDGENAVVHYTISGRDQHYFDINTKTGVVSTKLELK 2659

Qy        713 RVQLSPPQLAYIVVTRSMNTRFFLNGQNPPPGTIVDDVITLPERY--------DFYLVSQ 764

Db       2660 TKTKSHDDLTYTIVISAMDQ----GEQSLSSKAELTVILRPPELFPTFAYMANSHFAMSE 2715

Qy        765 QVRQG----TVSPTS----------YNVLYSSMGXS----PEK------MQKLTYKMCHL 800

Db       2716 DVRPGKMITKVSATSPKKGLVGKIRYAIAGGIMGDSLRVDPNSGLLSVGQDGLDYELTHL 2775

Qy        801 YYNW----SGTTRVPAVCQYAKKLATLVGTNLHSIPQNA--LEKKFY 841

Db       2776 YEIWIEAADGDT--PSL-----RSVTLITLNVTDANDNAPVMEQLIY 2815
```

```
RESULT 13
ANT1_ONCVO
ID   ANT1_ONCVO     STANDARD;      PRT;  2022 AA.
AC   P21249;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Major antigen.
GN   OVT1.
OS   Onchocerca volvulus.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC   Onchocercidae; Onchocerca.
OX   NCBI_TaxID=6282;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95287898; PubMed=7770081;
RA   Triteeraprapab S., Richie T.L., Tuan R.S., Shepley K.J., Dinman J.D.,
RA   Neubert T.A., Scott A.L.;
RT   "Molecular cloning of a gene expressed during early embryonic
RT   development in Onchocerca volvulus.";
RL   Mol. Biochem. Parasitol. 69:161-171(1995).
RN   [2]
RP   SEQUENCE OF 733-866 FROM N.A.
RX   MEDLINE=89127417; PubMed=2464764;
RA   Donelson J.E., Duke B.O.L., Moser D., Zeng W., Erondu N.E.,
RA   Lucius R., Renz A., Karam M., Flores G.Z.;
RT   "Construction of Onchocerca volvulus cDNA libraries and partial
RT   characterization of the cDNA for a major antigen.";
RL   Mol. Biochem. Parasitol. 31:241-250(1988).
CC   -!- FUNCTION: MAY BE A MYOFIBRILLAR PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U12681; AAA80009.1; -.
DR   EMBL; J03995; AAA29412.1; -.
DR   PIR; A54513; A54513.
KW   Antigen; Coiled coil.
FT   DOMAIN       74    120       COILED COIL (POTENTIAL).
FT   DOMAIN      151    251       COILED COIL (POTENTIAL).
FT   DOMAIN      327    384       COILED COIL (POTENTIAL).
FT   DOMAIN      417   1879       COILED COIL (POTENTIAL).
SQ   SEQUENCE   2022 AA;  237341 MW;  B7132AACF1520317 CRC64;

  Query Match             2.8%;  Score 121.5;  DB 1;  Length 2022;
  Best Local Similarity   20.2%;  Pred. No. 3.6;
  Matches  146;  Conservative  111;  Mismatches  266;  Indels  201;  Gaps   33;

Qy        160 EQEITVL------SGKSK-----LDIEYKISIKFVGFISCAE---PRFLQVLNLILRRSM 205

Db       1353 EKKETVLQTQKSPDGKSKERFDISDLDTNIQ-KLIGRIEKLELERNEYRDALDRIKKKSI 1411

Qy        206 KG---LN-LELVGRNLFDPRAKIEIREFKME--------LWPGYETSIRQHEKDILLGTE 253

Db       1412 ESHIKINKQETIFTNIEDQLVDVEEEKRTLEMRLASAKQLLRSQEEALKQRDE------E 1465

Qy        254 ITHKVMRTETIYDIMRRCSHNPARHQDEVRVNVL--------DLIVLTDYNNRTY----- 300

Db       1466 RSHMKLKIAK-FEMEARGKEAQLRQLNELVRNLRKDLETAQGDLGVLHDHEERWYAHKFH 1524

Qy        301 ---------------RINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLI 345

Db       1525 LESKLKDQESESQQIRLLVANF-ETERNSLNEKVRDLASRLQQTESKNADMKEDNDRL-- 1581

Qy        346 SKNRDKALKTNASELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAF 405

Db       1582 KKDLIKA-STNEAELRRTIDQNSRVVSDNQILKD--QLESAQND-----------LSNA 1626

Qy        406 NHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHF--- 462

Db       1627 NNRKQQLENELLVVRSELRDLKQRFSDNANRII----------------DLQRHLTDA 1668

Qy        463 -RDQRMLTTPSDGLDRWAVIAPQRN-SHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDR 520

Db       1669 ENDKKRLTNRLNSLEK--TVSQQRTIETEIRQQLSLALNERNTLQNDLRDLQRRLA---R 1723

Qy        521 TGTYVRAMDDCVRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLM 580

Db       1724 METEKKIMND----------------KYDELEK----IRASLIKRIELLDEEKRTME 1760

Qy        581 SIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGAL---IASMDLQQ 637

Db       1761 NILHETALQREA--------IESSLNAL-------ERENKELHRNCAQLQQQIAQLELEN 1805

Qy        638 NSTYFS-TVTECSAFDVLANTLWPMIAKALRQYQHEHRKLPSRI-------VFYRDGVSS 689

Db       1806 GNRLIQLTNKQREEYDKFAQNMRTEKIQIERIIENRERSLKSRINQLENQLNIMRDQLNS 1865

Qy        690 GSLKQLFEFEVKDIIEKLKTEYARVQLSPPQLAYIVVTRSM---NTRF----FLNGQNPP 742
```

```
Db       1866 ERRRRR---EISDKILSGEVNKLNVTLSGIPDAYEMYDRTLYTYNTYFGSPSFTVGSSSF 1922

Qy        743 PGTIVDD--------------------------VITLP--------ERYDFYLVSQQVRQ 768
                | ||                          :| |        :| | : |::  |
Db       1923 DPNITDDSKIILKHSDRLESSYAYGGGNRTSDTAVTAPTGSSSYHGQRSDTHTVNESERG 1982

Qy        769 GTVS 772
                | |
Db       1983 GNDS 1986


RESULT 14
SKI2_YEAST
ID   SKI2_YEAST     STANDARD;      PRT;  1287 AA.
AC   P35207; Q06047;
DT   01-FEB-1994 (Rel. 28, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Antiviral protein SKI2.
GN   SKI2 OR YLR398C OR L8084.17.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93309467; PubMed=8321235;
RA   Widner W.R., Wickner R.B.;
RT   "Evidence that the SKI antiviral system of Saccharomyces cerevisiae
RT   acts by blocking expression of viral mRNA.";
RL   Mol. Cell. Biol. 13:4331-4341(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313267; PubMed=9169871;
RA   Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansorge W.,
RA   Benes V., Bruckner M., Delius H., Dubois E., Dusterhoft A.,
RA   Entian K.D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA   Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Kotter P.,
RA   Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Mostl D.,
RA   Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA   Portetelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA   Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA   Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhasselt P.,
RA   Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
RA   Wedler H., Zimmermann F.K., Zollner A., Hani J., Hoheisel J.D.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RL   Nature 387:87-90(1997).
RN   [3]
RP   SEQUENCE OF 1-162 FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=95116323; PubMed=7816623;
RA   Lygerou Z., Conesa C., Lesage P., Swanson R.N., Ruet A., Carlson M.,
RA   Sentenac A., Seraphin B.;
RT   "The yeast BDF1 gene encodes a transcription factor involved in the
RT   expression of a broad class of genes including snRNAs.";
RL   Nucleic Acids Res. 22:5332-5340(1994).
CC   -!- FUNCTION: REPRESSES DSRNA VIRUS PROPAGATION BY SPECIFICALLY
CC       BLOCKING TRANSLATION OF VIRAL MRNAS, PERHAPS RECOGNIZING THE
CC       ABSENCE OF CAP OR POLY(A). ESSENTIAL FOR CELL GROWTH ONLY IN THE
CC       PRESENCE OF M1 REPLICON. SEEMS TO BE A HELICASE.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. SKI2 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L13469; AAA35049.1; -.
DR   EMBL; U19729; AAB82356.1; -.
DR   EMBL; Z18944; CAA79378.1; -.
DR   PIR; B48140; B48140.
DR   SGD; S0004390; SKI2.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   Antiviral; Translation regulation; RNA-binding; DNA-binding; Helicase;
KW   ATP-binding.
FT   NP_BIND     351    358       ATP (POTENTIAL).
FT   SITE        444    447       DEVH BOX.
FT   DOMAIN      556    577       RNA-BINDING RGG-BOX (BY SIMILARITY).
FT   DOMAIN      555    597       ARG/GLY-RICH.
FT   CONFLICT    326    326       W -> C (IN REF. 1).
FT   CONFLICT    759    760       QM -> L (IN REF. 1).
SQ   SEQUENCE   1287 AA;  146058 MW;  7CCD36CFC0DF8C32 CRC64;

  Query Match             2.7%;  Score 116.5;  DB 1;  Length 1287;
  Best Local Similarity   19.5%;  Pred. No. 4;
  Matches  137;  Conservative   90;  Mismatches  236;  Indels  239;  Gaps   37;

Qy          4 DQGRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRR- 62
              | |||           ||:||   |   :  ||      |       ||   |:    |:
Db        553 DNGRG---------GSTARGGRGGSNTRDGRGGRGNSTRGGANRGGSRGAGAIGSNKRKF 603

Qy         63 --EGGPPERKPWGDQYDYLNTRP-----VELVSKKGTDGVPVMLQ-TNFFRXKTKPEWRI 114
                : | | :| | :  :||  |      | : |||  :     |:  ||   | | :
Db        604 FTQDG-PSKKTWPEIVNYLRKRELLPMVVFVFSKKRCEEYADWLEGINFCNNKEKSQ--- 659

Qy        115 VHYHVE-----------FVPTIENPR--VRMGVLSNHANLL-----------GSGY---L 147
              :|  :|           :| |   |  :  |:  :|  ||                |:   |
Db        660 IHMFIEKSITRLKKEDRDLPQILKTRSLLERGIAVHHGGLLPIVKELIEILFSKGFIKVL 719

Qy        148 F-------------------------DGLQLFTTRKFEQEITVLSGKSKLD-------I 174
              |                         :||:  |  :| | :   :|:  ||        :
Db        720 FATETFAMGLNLPTRTVIFSSIRKHDGNGLRELTPGEFTQ-MAGRAGRRGLDSTGTVIVM 778

Qy        175 EYKISIKFVGF--ISCAEPRFLQ---------VLNLILRRSMKGLNLELVGRNLFDPRAK 223
              |   :    |  ::   |  ||          :|||:    :: | :| :.:  |   ||
Db        779 AYNSPLSIATFKEVTMGVPTRLQSQFRLTYNMILNLL---RIEALRVEEMIKYSFSENAK 835

Qy        224 IEIREFKMELWPGYETSIRQHEKDILLGTEITHKVMRTE--TI-YDIMRRCSHNPARHQD 280
                          ||   :||| |         ||:: |  || |     | ::  :  :
Db        836 -------------ETLQPEHEKQI--------KVLQEELQTIEYKSCEICDNDIEKFLE 873

Qy        281 ------EVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTFSCK-GRDISFVEYYLTKYN 333
                  |  ||::  :|                ::|     | || ::|
Db        874 LMLAYKEATVNLMQEMV----------------KSPSILHILKEGRLVAF--------- 907

Qy        334 IRIRDHNQPLLISKNRDKALKTNASELVVLIPELCRVTGL-NAEMRSNFQLMRAMSSYTR 392
                || |  | :        | :  : | :|    :   | | |        :|   | |
Db        908 ---RDPNDCLKLG----FVFKVSLKDAVCVIMTFTKPYKLPNGEPNHLIYFPKA-DGYRR 959

Qy        393 MN-PK-QRTD---------RLRAFNHRLQNTP------ESVKVLRDWNME---------- 425
              | || |:||          :     |    |      : |  | ::| |
Db        960 RNFPKFQKTDFYMEEVPVTAIEVITKRKFAAPLGKVIKKDVAALNEFNAETNNILDGKTL 1019

Qy        426 -----LDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDGLDRWAV 480
                  ::|  ::   :: : ||      |: : :  :  :|
Db       1020 KEAINIEKQGLKIHQILLDRTNIRDEIFKLKSIKCPNLSQH---------------- 1060

Qy        481 IAPQRNSHELRTLLDSLYRAASGMG----------LRIRSPQEFIIYDDRTGTYVRAMDD 530
              | |:  :| ::  :: ||   |              | :    |||  |       ::
Db       1061 IVPKFKAHVIKKKIEELYHLMSDQNLSLLPDYEKRLAVLKDTEFI--DQNHNVLLKGRVA 1118

Qy        531 C-VRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTL 571
              | : |  :|:|  :  ||          ::    | ::| |
Db       1119 CEINSGYELVLTELILDN----------FLGSFEPEEIVAL 1149
```

```
RESULT 15
ARS2_DROME
ID   ARS2_DROME     STANDARD;      PRT;   943 AA.
AC   Q9V9K7;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Arsenite-resistance protein 2 homolog.
GN   CG7843.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   REVISIONS, AND ALTERNATIVE SPLICING.
RC   STRAIN=Berkeley;
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celniker S.E.,
RA   Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,
RA   Harris N.L., Kronmiller B., Marshall B., Millburn G.H., Richter J.,
RA   Russo S., Searle S.M.J., Smith E., Shu S., Smutniak F.,
RA   Whitfield E.J., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; a long form (shown here) and a
CC       short form; are produced by alternative splicing.
CC   -!- SIMILARITY: BELONGS TO THE ARS2 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE003784; AAM68343.1; -.
DR   EMBL; AE003784; AAM68345.1; -.
DR   FlyBase; FBgn0033062; CG7843.
KW   Hypothetical protein; Alternative splicing.
FT   VARSPLIC     47     50       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   943 AA;  107221 MW;  0C1AF09E02E8AB0B CRC64;

  Query Match             2.6%;  Score 115.5;  DB 1;  Length 943;
  Best Local Similarity   19.5%;  Pred. No. 3;
  Matches  123;  Conservative   88;  Mismatches  208;  Indels  211;  Gaps   31;

Qy          9 RRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDP-RADPRIEASRERRALEEAPRREGGPP 67
              |:|  ::||||:|  |       |:        |  ||: : || ::|   :|| : : ||
Db        369 RKRNSSDDDSSSSESSSSSDEEKLKEKYDVEDGLRAEQKTEAEKDR---QEATKAKQGPQ 425

Qy         68 ERKPWGDQYDYLNTRPVELVSKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIEN 127
                |   |: :  || |  | ||                          :: : |   |:::
Db        426 SPKLDEDEGNE-NTEPKGLDSK-------------------------INTYEEIDNTLKS 459

Qy        128 PRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSK------LDIEYKISIK 181
              | :    : |  |  ||    ||          :: :|  || |      :|::
Db        460 PEISSNPIKNTDNGDGSKVEEDG----------EKPSV--GKDKVVETETIDLDK----- 502

Qy        182 FVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSI 241
                  :   :|| |   : |  |:
Db        503 ----VKDGQPRALHRTSSIFLRN------------------------------------- 521

Qy        242 RQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTYR 301
                    |   ||          :|   |:    |   :|| : | :|      | ||
Db        522 --------LAPSITRS--------EIEAVCN----RFSGYLRVAIADPLV----ERRWYR 557

Qy        302 INDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALK-----TN 356
              : |                ||::  |       | |:||     :::::  : ::     |
Db        558 RGWITF-----------MRDVNIKEICWGLNNQRLRDCEMGAIVNRDLSRRVRPANGITA 606

Qy        357 ASELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTPESV 416
                ::|    :||    || :    |:|           ||  ::  || |   :|   |
Db        607 HKQVVRSDIKLCAKIALNLD--EKFRLWAE-------GPKDDSNSARA-NESSENGSGST 656

Qy        417 KVLRDWNMELDKNVTEV--------QGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRML 468
                  |  | :|:|:          :  ::|              ||| |
Db        657 YGFNSQNPVL-QNITDYLIEEASAEEEELLGL-----------TGENKD----------- 693

Qy        469 TTPSDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAM 528
              |  : ::|           :| ::|| |        | :|      |:     |   |
Db        694 -TEGEPIER---------DEQLISVLDRLV-------LYLRIVHSVDYYNHCEYPYEDEM 736

Qy        529 DD-C----VRSDPKLILCLVPNDNAERYSSI---KKRGYVDRAVPTQVVTLKTTKNRSLM 580
              : |     |  | :    | |:: : |  |   | : :: : ||   :: :  ||
Db        737 PNRCGIIHARGPPPV---RVTNNDVQEYIKIYESKLQQFLTKTVP---LSDEEIKNLGAK 790

Qy        581 SIATKIA--IQLNC-KLGYTPWMIELPLSG 607
                |::   :| |  :|    |:   ||||
Db        791 DAETEVEKFVQANTQELAKDKWL--CPLSG 818


Search completed: July 22, 2003, 07:02:56
Job time : 13.8558 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 06:19:03 ; Search time 32.8527 Seconds
(without alignments)
5287.169 Million cell updates/sec

Title: US-09-873-737A-2
Perfect score: 4389
Sequence: 1 MADDQGRGRRRPXNEDDSST.....VGTNLHSIPQNALEKKFYYL 843

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4361	99.4	843	5	Q9VKM1	Q9vkm1 drosophila
2	1801.5	41.0	866	5	O76922	O76922 drosophila
3	1481.5	33.8	854	5	Q9GPA8	Q9gpa8 strongyloce
4	1424	32.4	861	4	Q96J94	Q96j94 homo sapien
5	1424	32.4	861	4	Q8TBY5	Q8tby5 homo sapien
6	1416.5	32.3	775	4	Q96JD5	Q96jd5 homo sapien
7	1412	32.2	861	4	O95404	O95404 homo sapien
8	1405.5	32.0	862	11	Q9JMB7	Q9jmb7 mus musculu
9	1332	30.3	858	13	Q8UVX0	Q8uvx0 brachydanio
10	1326.5	30.2	808	5	Q9GPA7	Q9gpa7 strongyloce
11	1182.5	26.9	971	11	Q9JMB6	Q9jmb6 mus musculu
12	1165	26.5	973	4	Q8TC59	Q8tc59 homo sapien
13	1087.5	24.8	824	5	P90786	P90786 caenorhabdi
14	953.5	21.7	722	5	Q17567	Q17567 caenorhabdi
15	875.5	19.9	580	11	Q99MV6	Q99mv6 mus musculu
16	823.5	18.8	421	5	Q9GPA6	Q9gpa6 strongyloce
17	788	18.0	530	4	Q96SW6	Q96sw6 homo sapien
18	646.5	14.7	781	5	Q9U5C9	Q9u5c9 paramecium
19	582	13.3	371	4	Q9NW28	Q9nw28 homo sapien
20	474	10.8	1000	5	Q20578	Q20578 caenorhabdi
21	465.5	10.6	891	5	O16720	O16720 caenorhabdi
22	458.5	10.4	950	5	Q9V3L2	Q9v3l2 drosophila
23	458.5	10.4	984	5	Q9V6V6	Q9v6v6 drosophila
24	449.5	10.2	924	4	Q9HCK5	Q9hck5 homo sapien
25	449	10.2	924	10	Q9ZVD5	Q9zvd5 arabidopsis
26	448	10.2	860	4	Q9H9G7	Q9h9g7 homo sapien
27	445.5	10.2	851	4	Q8TCZ5	Q8tcz5 homo sapien
28	444	10.1	904	10	Q9SDG8	Q9sdg8 oryza sativ
29	433	9.9	1035	5	Q21079	Q21079 caenorhabdi
30	425.5	9.7	1037	5	Q9TW94	Q9tw94 caenorhabdi
31	410.5	9.4	962	5	Q9VUQ6	Q9vuq6 drosophila
32	404	9.2	990	10	Q9C793	Q9c793 arabidopsis
33	395.5	9.0	1123	10	Q9LP83	Q9lp83 arabidopsis
34	395	9.0	640	4	Q9H1U6	Q9h1u6 homo sapien
35	386	8.8	585	4	Q8WV58	Q8wv58 homo sapien
36	383.5	8.7	694	5	Q8T3N2	Q8t3n2 drosophila
37	380.5	8.7	189	5	Q95PE5	Q95pe5 drosophila
38	368	8.4	601	5	Q960T1	Q960t1 drosophila
39	360.5	8.2	1014	10	Q9SHF3	Q9shf3 arabidopsis
40	357.5	8.1	1194	10	Q9SHF2	Q9shf2 arabidopsis
41	353.5	8.1	887	10	O48771	O48771 arabidopsis
42	351.5	8.0	898	5	Q9XVI3	Q9xvi3 caenorhabdi
43	330	7.5	377	4	Q96ID1	Q96id1 homo sapien
44	330	7.5	377	11	Q8R3Q7	Q8r3q7 mus musculu
45	315	7.2	938	3	Q9P8T1	Q9p8t1 neurospora

ALIGNMENTS

RESULT 1
Q9VKM1
ID Q9VKM1 PRELIMINARY; PRT; 843 AA.
AC Q9VKM1; O96674; O96675;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Piwi protein.
GN PIWI OR CG6122.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=OREGON-R, AND CANTON-S;
RX MEDLINE=99069219; PubMed=9851978;
RA Cox D.N., Chao A., Baker J., Chang L., Qiao D., Lin H.;
RT "A novel class of evolutionarily conserved genes defined by piwi are
RT essential for stem cell self-renewal.";
RL Genes Dev. 12:3715-3727(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

```
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: MEDIATES A SOMATIC SIGNALING MECHANISM REQUIRED FOR THE
CC       ASYMMETRIC DIVISION OF GERM-LINE STEM CELLS TO PRODUCE AND
CC       MAINTAIN A DAUGHTER GERM-LINE STEM CELL. IT IS NOT ESSENTIAL FOR
CC       THE FURTHER DIFFERENTIATION OF THE COMMITTED DAUGHTER CELL.
CC   -!- TISSUE SPECIFICITY: EXPRESSED SOMATICALLY IN THE OVARIOLE TERMINAL
CC       FILAMENT AND EPITHELIAL SHEATH CELLS. EXPRESSED IN THE GERMARIUM
CC       AT LOW LEVELS DURING OOGENESIS STAGES 1-6, AT A LOWER LEVEL DURING
CC       STAGES 7-9, STRONGLY AT STAGE 10 AND EVENTUALLY ACCUMULATING IN
CC       EARLY EMBRYOS.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED BOTH MATERNALLY AND ZYGOTICALLY.
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
DR   EMBL; AF104354; AAD08704.1; -.
DR   EMBL; AF104355; AAD08705.1; -.
DR   EMBL; AE003630; AAF53043.1; -.
DR   FlyBase; FBgn0004872; piwi.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
KW   Developmental protein.
FT   CONFLICT     67     67       T -> P (IN REF. 1; AAD08704).
FT   CONFLICT     84     84       A -> V (IN REF. 1; AAD08704).
FT   CONFLICT    122    124       EPS -> VPT (IN REF. 1; AAD08704).
FT   CONFLICT    576    577       NR -> KPY (IN REF. 1; AAD08705).
SQ   SEQUENCE   843 AA;  97177 MW;  9B4A95E688E9D9B7 CRC64;

  Query Match             99.4%;  Score 4361;  DB 5;  Length 843;
  Best Local Similarity   99.2%;  Pred. No. 7.3e-306;
  Matches  836;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 MADDQGRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAP 60
              ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MADDQGRGRRRPLNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAP 60

Qy         61 RREGGPPERKPWGDQYDYLNTRPVELVSKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVE 120
              |||||| |||||||||||||| |||||||||||||||||||| ||||||||||||||||
Db         61 RREGGPTERKPWGDQYDYLNTRPAELVSKKGTDGVPVMLQTNFFRLKTKPEWRIVHYHVE 120

Qy        121 FVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISI 180
              | |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 FEPSIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISI 180

Qy        181 KFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETS 240

Qy        241 IRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTY 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        241 IRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTY 300

Qy        301 RINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASEL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALKTNASEL 360

Qy        361 VVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLR 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLR 420

Qy        421 DWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDGLDRWAV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHFRDQRMLTTPSDGLDRWAV 480

Qy        481 IAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCVRSDPKLIL 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 IAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCVRSDPKLIL 540

Qy        541 CLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLGYTPWM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKNRSLMSIATKIAIQLNCKLGYTPWM 600

Qy        601 IELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWP 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWP 660

Qy        661 MIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLSPPQ 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 MIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLSPPQ 720

Qy        721 LAYIVVTRSMNTRFFLNGQNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLY 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 LAYIVVTRSMNTRFFLNGQNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLY 780

Qy        781 SSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKF 840
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 SSMGLSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKF 840

Qy        841 YYL 843
              |||
Db        841 YYL 843


RESULT 2
O76922
ID   O76922      PRELIMINARY;      PRT;   866 AA.
AC   O76922;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   AUB protein.
GN   AUB OR STING OR CG6137.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
```

```
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON S;
RX   MEDLINE=96262192; PubMed=8676882;
RA   Schmidt A., Hollmann M., Schaefer U.;
RT   "A newly identified Minute locus, M(2)32D, encodes the ribosomal
RT   protein L9 in Drosophila melanogaster.";
RL   Mol. Gen. Genet. 251:381-387(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON S;
RA   Schmidt A., Palumbo G., Bozzetti M.P., Tritto P., Pimpinelli S.,
RA   Schaefer U.;
RT   "Genetic and molecular characterization of sting, a gene involved in
RT   crystal formation and meiotic drive in Drosophila melanogaster.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Tsang G., Brokstein P., Frise E., Harvey D., Evans-Holm M.,
RA   Lewis S.E., Suh C., Rubin G.M.;
RT   "Full Length Drosophila melanogaster cDNA sequence.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003630; AAF53046.1; -.
DR   EMBL; X94613; CAA64320.1; -.
DR   EMBL; AF145680; AAD38655.1; -.
DR   FlyBase; FBgn0000146; aub.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   866 AA;  98558 MW;  8F84CF1B0B89103B CRC64;

  Query Match             41.0%;  Score 1801.5;  DB 5;  Length 866;
  Best Local Similarity   42.4%;  Pred. No. 3.5e-121;
  Matches  374;  Conservative  159;  Mismatches  279;  Indels   71;  Gaps   13;

Qy          5 QGRGR-RRPXN------------EDDSSTSRGS---------GD---GPRVK------V 32
              :|||| |:| |             : | | | |:          ||   || ::      |
Db         11 RGRGRGRKPNNVEANRGFAPSLGQKSDPSHSEGNQASGGNGGGGDAQVGPSIEKSSLSAV 70

Qy         33 FRGSSSGDPRADPRIEASRERRALEEAPRREGGPPERKPWGDQYDYLNTRPVELVSKKGT 92
                | ||||     : | || :                          | : :||  : ||||
Db         71 QMHKSEGDPRG-----SVRGRRLIT------------------DLVYSRPPGMTSKKGV 106

Qy         93 DGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMGVLSNHANLLGSGYLFDGLQ 152
                |  : :| |:|:   :| | |  | |:| | :|  |:|   |  |  :|| ||:|||
Db        107 VGTHITVQANYFKVLKRPNWTIYQYRVDFTPDVEATRLRRSFLYEHKGILG-GYIFDGTN 165

Qy        153 LFTTRKFE-----QEITVLSGKSKLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKG 207
              :|   :|:       :  |  ||:     :| || || :   :    |||||||||:|:|
Db        166 MFCINQFKAVQDSPYVLELVTKSRAGENIEIKIKAVGSVQSTDAEQFQVLNLILRRAMEG 225

Qy        208 LNLELVGRNLFDPRAKIEIREFKMELWPGYETSIRQHEKDILLGTEITHKVMRTETIYDI 267
              |:|:|| |  :||:||| :  |:|:|||||:|||||| |||| :|| |||||||:|:|
Db        226 LDLKLVSRYYYDPQAKINLENFRMQLWPGYQTSIRQHENDILLCSEICHKVMRTETLYNI 285

Qy        268 MRRCSHNPARHQDEVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTFSCKGRDISFVEY 327
              :    :   :|   :  |: :::|||||:||||:|||  ||   |     :||:|:|
Db        286 LSDAIRDSDDYQSTFKRAVMGMVILTDYNNKTYRIDDVDFQSTPLCKFKTNDGEISYVDY 345

Qy        328 YLTKYNIRIRDHNQPLLISKNRDKALKTNASELVVLIPELCRVTGLNAEMRSNFQLMRAM 387
              |  :||| |||  |||::|:  || ::    : :::|||| | ||:   ||::|: :|||
Db        346 YKKRYNIIIRDLKQPLVMSRPTDKNIRGGNDQAIMIIPELARATGMTDAMRADFRTLRAM 405

Qy        388 SSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHN 447
              | :||:|| :| :||| || ||::  :||: |: ||:|||  : |:  |::  : |:| |
Db        406 SEHTRLNPDRRIERLRMFNKRLKSCKQSVETLKSWNIELDSALVEIPARVLPPEKILFGN 465

Qy        448 GKV-PAGENADWQRHFRDQRMLTTPSDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGL 506
               |:       |||   ||   |       ::|| || | ||  | :  :    | || | :
Db        466 QKIFVCDARADWTNEFRTCSMFKNVH--INRWYVITPSRNLRETQEFVQMCIRTASSMKM 523

Qy        507 RIRSPQEFIIYDDRTGTYVRAMDDCVRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPT 566
              | :|    | ||| ||| :|:|:   :||:::: :: : | |:|| |||  ||| ||:
Db        524 NICNPIYEEIPDDRNGTYSQAIDNAAANDPQIVMVVMRSPNEEKYSCIKKRTCVDRPVPS 583

Qy        567 QVVTLKT-----TKNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRD 621
              ||||||       |   |||||||: ||:| ||   || : :|| ||||:|||:  | ::
Db        584 QVVTLKVIAPRQQKPTGLMSIATKVVIQMNAKLMGAPWQVVIPLHGLMTVGFDVCHSPKN 643

Qy        622 RKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIV 681
              : ::||| :|:|| :::  ||||| |       |:  :   :| ||| || :|| || ||:
Db        644 KNKSYGAFVATMDQKESFRYFSTVNEHIKGQELSEQMSVNMACALRSYQEQHRSLPERIL 703

Qy        682 FYRDGVSSGSLKQLFEFEVKDIIEKLKTEYARV-QLSPPQLAYIVVTRSMNTRFFLNGQN 740
              |:|||  | | |:   || : ::|  |    :    :: :|:|:: :|:|:|   :|
Db        704 FFRDGVGDGQLYQVVNSEVNTLKDRLDEIYKSAGKQEGCRMTFIIVSKRINSRYFTGHRN 763

Qy        741 PPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHL 800
              | |||:|||||||||||:|||| || |||||||||:  :|| : :|:| |:||| |:
Db        764 PVPGTVVDDVITLPERYDFFLVSQAVRIGTVSPTSYNVISDNMGLNADKLQMLSYKMTHM 823

Qy        801 YYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKFYYL 843
              |||:||| |||||| || ||| ||  :::  |   |: : |:|
Db        824 YYNYSGTIRVPAVCHYAHKLAFLVAESINRAPSAGLQNQLYFL 866


RESULT 3
Q9GPA8
ID   Q9GPA8      PRELIMINARY;      PRT;   854 AA.
AC   Q9GPA8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Seawi.
OS   Strongylocentrotus purpuratus (Purple sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Strongylocentrotus.
OX   NCBI_TaxID=7668;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rodriguez A.J., Bonder E.M.;
RT   "Seawi - Cloning and Molecular Characterization of a Sea Urchin
RT   Homolog of Piwi.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY014899; AAG42533.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
```

```
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   854 AA;  96722 MW;  509A1D39C0D1922C CRC64;

  Query Match           33.8%;  Score 1481.5;  DB 5;  Length 854;
  Best Local Similarity  36.5%;  Pred. No. 4.1e-98;
  Matches  314;  Conservative  176;  Mismatches  313;  Indels   57;  Gaps   16;

Qy          6 GRGRRRPXNEDDSSTSRGSGDGPRVKVFRG----------SSSGDPRADPRIEASRERR 54
              |||| |       :|. | |  |: :  ||            :|||: |    ::  | ||
Db         30 GRGRSR-----GGTTQPGPGAVPKQEP-RGELTSQIAKLAIGASGEGRVGGAVQDRRRRR 83

Qy         55 -ALEEAPRREGGPPERKPWGDQYDYLNTRPVELVSKKGTDGVPVMLQTNFFRXKTKPEWR 113
               :|  |                    |||   | |:   |  : |  | |: ||||:|:
Db         84 DEIELEP------------------TTRPENFV-KQAIAGDKIALIANGFKLKTKPDWQ 123

Qy        114 IVHYHVEFVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLD 173
              :  | |:|  | |  ||| |  :|  |: ||| |    |   :::  |  :::| || : |
Db        124 LYQYRVDFEPEILNPRARFALLKGHSALLGKGLTLDMDTMYSLFKLAEKVTNLSAERKDG 183

Qy        174 IEYKISIKFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMEL 233
                | :  |  ::   |  | : |:: || :| :|:| |||| :|| | |:|::  ::|
Db        184 SNVNICVTHVATLNPMAPNTLHLYNVLFRRCLKMINMEQVGRNYYDPTAAIDIKQHGLQL 243

Qy        234 WPGYETSIRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNP-ARHQDEVRVNVLDLIVL 292
              |||: ||| |:| |::| ::|:|||:||:|::::|    :    | ::|:   ::  |||
Db        244 WPGFVTSILQYEYDVMLLSDISHKVLRTQTVWEVMNDLFNKARGRFKEEITKLMIGQIVL 303

Qy        293 TDYNNRTYRINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKA 352
              | |||:|||::|:||  ||  ||  :   :|:|:|:    |  | | |||:|:|: : :
Db        304 TKYNNKTYRVDDIDFDTTPADTFETRSGPVSYVDYFKKSYERVIHDVNQPMLVSRPKKRE 363

Qy        353 LKTNASELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNT 412
              |       | |||| :|||: :||::| :|: |  :|||: |: |   |  |  :| :
Db        364 EKKGVGP-AYLPPELCFLTGLSDDMRADFNVMKDMGQHTRVGPQDRCRTLSGFIKKLSSN 422

Qy        413 PESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKV----PAGENADWQRHFRDQRML 468
              |    |  | || ||   :: ||::  : : |  ||     |:  |||| |  |   :
Db        423 EEVKTYLDSWGMEFDKEQVKLTGRVLPPEKL-FQRGKQFSYNPS--NADWSRDTRGNAL- 478

Qy        469 TTPSDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAM 528
              | :  |: | :   :|:::  :  : || | |: ||: :| |:   : |||| || |::
Db        479 -TDAKILNNWKIFYTRRDANRGQDFIKSLVRVANPMGMNVRGPEIVELPDDRTETYTRSL 537

Qy        529 DDCVRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKT-TKNRSLMSIATKIA 587
              :  | :::: ::| :  :|| :|||    |    |:||:  :| :| : |||:|||||
Db        538 QAQIAQDTQIVVVILPTNRKDRYDAIKKTCVVTHPCPSQVIVSRTLSKQQMLMSVATKIA 597

Qy        588 IQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTE 647
              :|:|||:|   | :|:||| || || |    :  : |:    :|||:  | | :  |   :
Db        598 MQMNCKMGGDLWRVEIPLSNLMIIGIDSYHDSLTKGRSVLGFVASMNKSQTSFFSSCAFQ 657

Qy        648 CSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKL 707
              :  :  || |  ::  ||::|   : | | ||: :||||    :  : ::|:| | : |
Db        658 HAQGEFGAN-LSTLMNNALKRYYQINEKFPERIIIFRDGVGDSQVNLVVDYELKQIKDTL 716

Qy        708 KTEYARVQLSPPQLAYIVVTRSMNTRFFLNGQ----NPPPGTIVDDVITLPERYDFYLVS 763
              |   | :  :|| :|| : :| ||| | :    ||||||::|||:| |  ||::::|
Db        717 DKVYP--QGTVHKLAVVVVKKRINNRFFANLRGGLSNPPPGTVIDDVVTKPHLYDYFIIS 774

Qy        764 QQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATL 823
              | ||||:||||||||:| : |  |: ||:||||: |||:|| || |||| | || ||| |
Db        775 QSVRQGSVSPTSYNVVYDTTGLKPDHMQRLTYKLTHLYFNWPGTVRVPAPCMYAHKLAFL 834

Qy        824 VGTNLHSIPQNALEKKFYYL 843
              :| ::| :| | |    :||
Db        835 IGQSVHEVPNNRLADTLFYL 854


RESULT 4
Q96J94
ID   Q96J94      PRELIMINARY;      PRT;   861 AA.
AC   Q96J94;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   PIWI protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Sha J.H.;
RT   "Cloning and identification of human piwi protein related to testis
RT   development.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF387507; AAK69348.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   861 AA;  98603 MW;  58D7F6C7321DEFA4 CRC64;

  Query Match           32.4%;  Score 1424;  DB 4;  Length 861;
  Best Local Similarity  36.0%;  Pred. No. 5.9e-94;
  Matches  310;  Conservative  174;  Mismatches  308;  Indels   68;  Gaps   20;

Qy          6 GRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG 65
              ||||:|        |: |:     :::  |                          :|    | |
Db         48 GRGRQR-------GTAGGTAKSQGLQISAG--------------------FQELSLAERG 80

Qy         66 PPERKPWGDQYDY-LNTRP-VELV--SKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEF 121
                 |    | :|  :|||  :: |  || |: |: | | || ||  ::|:| :  ||:::
Db         81 GRRR----DFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDY 136

Qy        122 VPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIK 181
               | :|  |:|  :|  | :|:|  : |||  ||  :: :|::| :  |::   : :|:|
Db        137 NPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITIT 196

Qy        182 FVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSI 241
                  :    |  ||  |:| || :| :||: :||| ::|   |:|   :: :|||: |||
Db        197 LTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSI 256

Qy        242 RQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRT 299
               |:|  |:| |:::|||:|:||: | |    |    |  |::|   :: |:||| |||:|
Db        257 LQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKELIGLVVLTKYNNKT 316

Qy        300 YRINDVDFGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNA 357
              ||::|:|: | |||||    | ::||:||| :||  | |  ||:|:|:  | :
Db        317 YRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTL 376

Qy        358 SELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPE 414
                :|||||| :|||  :||::| :|: :: :||: |:||     ||  : |:  |
Db        377 PGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQ- 435

Qy        415 SVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTP 471
              : |||| :  | |:    |||:  : |  | |       |   ||| :  |   :::
Db        436 --RELRDWGLSFDSNLLSFSGRILQTEKI--HQGGKTFDYNPQFADWSKETRGAPLISVK 491

Qy        472 SDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDC 531
                | | :|  :||     :|: :|::    ||:::|     |  |||| |:| :
Db        492 P--LDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRK-AIMIEVDDRTEAYLRVLQQK 548

Qy        532 VRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQL 590
              | :| ::::||: ::  ::| :|||    |   |:| |  :|  | :::|:||||||:|:
Db        549 VTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQM 608

Qy        591 NCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSA 650
              |||:|   | :::||  :| :|  |         :|:    :|||:: :  : :||
Db        609 NCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASIN-EGMTRWFSRCIFQDR 667

Qy        651 FDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKT- 709
              | : |   :  ||| :   :  :||||: |||||  | || |  :||   :: ||:
Db        668 GQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSI 727
```

```
Qy        710 --EYARVQLSPPQLAYIVVTRSMNTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVS 763
                  |       |:|  ||| : :|||||       ||| |||::|  :| || |||::||
Db        728 GRGY------NPRLTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVS 781

Qy        764 QQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATL 823
              | || |:|||| |||:| : |  |: :|:||||:||:|||| |  |||| |||| ||| |
Db        782 QAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFL 841

Qy        824 VGTNLHSIPQNALEKKFYYL 843
              || ::|  |  :|  : |||
Db        842 VGQSIHREPNLSLSNRLYYL 861


RESULT 5
Q8TBY5
ID   Q8TBY5      PRELIMINARY;      PRT;   861 AA.
AC   Q8TBY5;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Piwi-like 1 (Drosophila).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Strausberg R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC028581; AAH28581.1; -.
SQ   SEQUENCE   861 AA;  98545 MW;  D33376EEDEDF743A CRC64;

  Query Match             32.4%;  Score 1424;  DB 4;  Length 861;
  Best Local Similarity   35.9%;  Pred. No. 5.9e-94;
  Matches  309;  Conservative  176;  Mismatches  307;  Indels   68;  Gaps   20;

Qy          6 GRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG 65
              ||||:|        |: |:     :::  |                  :|    | |
Db         48 GRGRQR-------GTAGGTAKSQGLQISAG-------------------FQELSLAERG 80

Qy         66 PPERKPWGDQYDY-LNTRP-VELV--SKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEF 121
                 |    | :|  :|||  :: |  || |: |: | | || ||  ::|:| :  ||::: 
Db         81 GRRR----DFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDY 136

Qy        122 VPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIK 181
              | :|  |:|  :|  | :|:|  : |||  ||  :: :|:::| :  |::   : :|:| 
Db        137 NPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITIT 196

Qy        182 FVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSI 241
                :    |  ||  |:| || :| :||: :||| ::|   |:|   :: :|||: |||
Db        197 LTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSI 256

Qy        242 RQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRT 299
              |:|  |:| |:::|||:|:||: | |    |    |  |::|   :: |:||| |||:|
Db        257 LQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKELIGLVVLTKYNNKT 316

Qy        300 YRINDVDFGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNA 357
              ||:|:|: | |||||    | ::||:||| :||  | |  ||:|:|:  | :
Db        317 YRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTL 376

Qy        358 SELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPE 414
                :|||||| :|||  :||::| :|: :: :||: |:||     ||  : |:  |
Db        377 PGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQ- 435

Qy        415 SVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTP 471
              : |||| :  | |:    |||:  : |  | |       |   ||| : |   :::
Db        436 --RELRDWGLSFDSNLLSFSGRILQTEKI--HQGGKTFDYNPQFADWSKETRGAPLISV- 490

Qy        472 SDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDC 531
              : || | :|  :||     :|: :|::    ||:::::     |  ||||  |:| :
Db        491 -NPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMKK-AIMIEVDDRTEAYLRVLQQK 548

Qy        532 VRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQL 590
              | :| ::::||: ::  ::| :|||     |   |:| |  :|  | :::|:|||||:|:
Db        549 VTADTQIVVCLLSSNRKDKYDAIKKYPCTDCPTPSQCVVARTLGKQQTVMAIATKIALQM 608

Qy        591 NCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSA 650
              |||:|   | :::||  :| :| |         :|:    :|||: :  : :||
Db        609 NCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASIN-EGMTRWFSRCIFQDR 667

Qy        651 FDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKT- 709
                | : |   :  ||| :   :  :||||: |||||  | || |  :||   :: ||:
Db        668 GQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSI 727

Qy        710 --EYARVQLSPPQLAYIVVTRSMNTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVS 763
                  |       |:|  ||| : :|||||       ||| |||::|  :| || |||::||
Db        728 GRGY------NPRLTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVS 781

Qy        764 QQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATL 823
              | || |:|||| |||:| : |  |: :|:||||:||:|||| |  |||| |||| ||| |
Db        782 QAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFL 841

Qy        824 VGTNLHSIPQNALEKKFYYL 843
              || ::|  |  :|  : |||
Db        842 VGQSIHREPNLSLSNRLYYL 861


RESULT 6
Q96JD5
ID   Q96JD5      PRELIMINARY;      PRT;   775 AA.
AC   Q96JD5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   HIWI.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Sharma A.K., Nelson M.C., Brandt J.E., Wessman M., Muhmud N.,
RA   Weller K.P., Hoffman R.;
RT   "Human CD34+ Stem Cells Express the hiwi Gene, a Human Homolog of the
RT   Drosophila Gene piwi.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF264004; AAK92281.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   775 AA;  89484 MW;  DF169A2E9EAFD916 CRC64;

  Query Match             32.3%;  Score 1416.5;  DB 4;  Length 775;
  Best Local Similarity   37.7%;  Pred. No. 1.7e-93;
  Matches  292;  Conservative  163;  Mismatches  286;  Indels   33;  Gaps   14;

Qy         88 SKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMGVLSNHANLLGSGYL 147
              || |: |: | | || ||  ::|:| :  ||::: | :|  |:|  :|  | :|:|  :
Db         17 SKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHA 76

Qy        148 FDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKG 207
              |||  ||  :: :|::| :  |::   : :|:|     :    | ||  |:| || :|
Db         77 FDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKI 136

Qy        208 LNLELVGRNLFDPRAKIEIREFKMELWPGYETSIRQHEKDILLGTEITHKVMRTETIYDI 267
              :||: :||| ::|   |:|   :: :|||: ||| |:|  |:| |:::|||:|:||: |
Db        137 MNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDF 196

Qy        268 MRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTF-SCKGRDISF 324
              |    |   |  |::|   :: |:||| |||:|||::|:|: | |||||    | ::||
Db        197 MFNFYHQTEEHKFQEQVSKELIGLVVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSF 256

Qy        325 VEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNASELVVLIPELCRVTGLNAEMRSNFQL 383
```

```
                :|||  :||  | |  ||:|:|:  | :          :|||||| :|||  :||::| :
Db        257 LEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNV 316

Qy        384 MRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQ 440
              |: :: :||: |:||     ||  : |:  |     : |||| :  | |:    |||:
Db        317 MKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQ---RELRDWGLSFDSNLLSFSGRILQT 373

Qy        441 QNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTPSDGLDRWAVIAPQRNSHELRTLLDSL 497
              : |  | |       |   ||| :  |   :::     || | :|  :||     :|: :|
Db        374 EKI--HQGGKTFDYNPQFADWSKETRGAPLISVKP--LDNWLLIYTRRNYEAANSLIQNL 429

Qy        498 YRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCVRSDPKLILCLVPNDNAERYSSIKKR 557
              ::    ||:::|     |  |||| |:| :   | :| ::::||: ::  ::| :|||
Db        430 FKVTPAMGMQMRK-AIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKY 488

Qy        558 GYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIA 616
                |  |:| |  :|  | :::|:|||||||:|:|||:|  | :::||  :| :| |
Db        489 LCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCY 548

Qy        617 KSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRKL 676
                  :|:    :||:: :  : :||          | : |   :  ||| :   :  :
Db        549 HDMTAGRRSIAGFVASIN-EGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYM 607

Qy        677 PSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKT---EYARVQLSPPQLAYIVVTRSMNTR 733
              ||||: |||||  | || |  :||   :: ||:    |       |:|  ||| : :|||
Db        608 PSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGY------NPRLTVIVVKKRVNTR 661

Qy        734 FFLNG----QNPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEK 789
              ||       ||| |||::|  :| || |||::||| || |:|||| |||:| : |  |:
Db        662 FFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDH 721

Qy        790 MQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKFYYL 843
              :|:||||:||:|||| |  |||| |||| ||| ||| ::|  |  :|  : |||
Db        722 IQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 775
```

RESULT 7
O95404
ID O95404 PRELIMINARY; PRT; 861 AA.
AC O95404;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE HIWI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE=99069219; PubMed=9851978;
RA Cox D.N., Chao A., Baker J., Chang L., Qiao D., Lin H.;
RT "A novel class of evolutionarily conserved genes defined by piwi are
RT essential for stem cell self-renewal.";
RL Genes Dev. 12:3715-3727(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Qiao D., Zeeman A.-M., Deng W., Looijenga L.H.J., Lin H.;
RT "Molecular characterization of hiwi, a human member of the piwi stem
RT cell gene family whose overexpression is correlated to seminomas.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF104260; AAC97371.2; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
SQ SEQUENCE 861 AA; 98530 MW; 43D7F60E99D997B7 CRC64;

Query Match 32.2%; Score 1412; DB 4; Length 861;
Best Local Similarity 35.9%; Pred. No. 4.3e-93;
Matches 309; Conservative 173; Mismatches 310; Indels 68; Gaps 20;

```
Qy          6 GRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG 65
              ||||:|        |: |:     :::  |                        :|    | |
Db         48 GRGRQR-------GTAGGTAKSQGLQISAG-------------------FQELSLAERG 80

Qy         66 PPERKPWGDQYDY-LNTRP-VELV--SKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEF 121
                 |    | :|  :|||  :: |  || |: |: | | || ||  ::|:| :  ||:::
Db         81 GRRR----DFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDY 136

Qy        122 VPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIK 181
               | :|  |:|  :|  | :|:|  : ||| ||  :: :|::| :  |::  : :|:|
Db        137 NPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITIT 196

Qy        182 FVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSI 241
                  :    |  ||  |:| || :| :||: :||| ::|   |:|   :: :|||: |||
Db        197 LTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSI 256

Qy        242 RQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRT 299
              |:|  |:| |:::|||:|:||: | |    |    |  |::|   :: |:||| || :|
Db        257 LQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKELIGLVVLTKYNIKT 316

Qy        300 YRINDVDFGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNA 357
              ||::|:|: | |||||     |  :|||:|||  :||  | |  ||:|:|:  | :
Db        317 YRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTL 376

Qy        358 SELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPE 414
                :|||||| :|||  :||::| :|: :: :||: |:||     ||  : |:  |
Db        377 PGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQ- 435

Qy        415 SVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTP 471
                : |||| :  | |:     |||:  : |  | |       |   ||| :  |   :::
Db        436 --RELRDWGLSFDSNLLSFSGRILQTEKI--HQGGKTFDYNPQFADWSKETRGAPLISVK 491

Qy        472 SDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDC 531
                || | :|  :||     :|: :|::    ||:::|     |  ||||  |:| :
Db        492 P--LDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRK-AIMIEVDDRTEAYLRVLQQK 548

Qy        532 VRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQL 590
              | :| ::::||: ::  ::| :|||     |   |:| |  :|  | :::|:|||||:|:
Db        549 VTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQM 608

Qy        591 NCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSA 650
              |||:|   | :::||  :| :| |         :|:    :|||:: :  : :||
Db        609 NCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASIN-EGMTRWFSRCIFQDR 667

Qy        651 FDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKT- 709
                | : |   :  ||| :   :  :||||: |||||  | || |  :||   :: ||:
Db        668 GQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSI 727

Qy        710 --EYARVQLSPPQLAYIVVTRSMNTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVS 763
                |       |:|  ||| : :|||||       ||| |||::|  :| || |||::||
Db        728 GRGY------NPRLTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVS 781

Qy        764 QQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATL 823
              | || |:|||| |||:| : |  |: :|:||||:||:|||| |  |||| |||| ||| |
Db        782 QAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFL 841

Qy        824 VGTNLHSIPQNALEKKFYYL 843
              || ::|  |  :|  : |||
Db        842 VGQSIHREPNLSLSNRLYYL 861
```

RESULT 8
Q9JMB7
ID Q9JMB7 PRELIMINARY; PRT; 862 AA.
AC Q9JMB7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE MIWI (piwi).
GN PIWIL1 OR MIWI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Miyagawa S.K., Kimura T., Nakano T.;
RT   "Molecular Cloning and characterization of Piwi family genes.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Deng W., Lin H.;
RT   "Miwi, a murine homolog of piwi, encodes a cytoplasmic protein
RT   essential for spermatogenesis.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB032604; BAA93705.1; -.
DR   EMBL; AF438405; AAL31014.1; -.
DR   MGD; MGI:1928897; Piwil1.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   862 AA;  98574 MW;  45588D13284CCC4C CRC64;

  Query Match           32.0%;  Score 1405.5;  DB 11;  Length 862;
  Best Local Similarity  36.1%;  Pred. No. 1.3e-92;
  Matches  309;  Conservative  174;  Mismatches  314;  Indels   59;  Gaps   19;

Qy          6 GRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG 65
              ||||:|      :| |:                    : :| |  :  :| |     ||
Db         48 GRGRQRGMVVGATSKSQ--------------------ELQISAGFQELSLAE----RGG 82

Qy         66 PPERKPWGDQYDYLNTRP-VELV--SKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFV 122
               |: : |    :|||  :: |  || |: |: | | || ||  ::|:| :  ||:::
Db         83 --RRRDFHDL--GVNTRQNLDHVKESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYN 138

Qy        123 PTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIKF 182
              | :|  |:|  :|  | :|:|  : |||  ||  :: : ::| :  :::     :|:|
Db        139 PLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVTEVFSQTRNGEHVRITITL 198

Qy        183 VGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSIR 242
                :    |  ||  |:| || :| :||: :||| ::|   |:|   :: :|||: |||
Db        199 TNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSIL 258

Qy        243 QHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRTY 300
              |:| :|:| |:::|||:|:||: | |          |  |::|   :: ||||| |||:||
Db        259 QYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTY 318

Qy        301 RINDVDFGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNAS 358
              |::|:|: | |||||    | ::||:||| :||  | |  ||:|:|:  | :
Db        319 RVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLP 378

Qy        359 ELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPES 415
                :|||||| :||| :||::| :|: :: :||: |:||     ||  : |:  |
Db        379 GPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQ-- 436

Qy        416 VKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTPS 472
               : |||| :  | |:    |||:  : |  | |       |   ||| :  |   :::
Db        437 -RELRDWGLSFDSNLLSFSGRILQSEKI--HQGGKTFDYNPQFADWSKETRGAPLISVKP 493

Qy        473 DGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCV 532
              || | :|  :||     :|: :|::    ||::::      |  ||||  |:||:   |
Db        494 --LDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKK-AIMIEVDDRTEAYLRALQQKV 550

Qy        533 RSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQLN 591
              || ::::||: ::  ::| :|||    |  |:| |  :|  | :::|:|||||:|:|
Db        551 TSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMN 610

Qy        592 CKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAF 651
              ||:|   | :::||   | :| |    |   :|:    :|||: :  : :||
Db        611 CKMGGELWRVDMPLKLAMIVGIDCYHDTTAGRRSIAGFVASIN-EGMTRWFSRCVFQDRG 669

Qy        652 DVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEY 711
                | : |   :  ||| :   :  :|||:: |||||  | || |  :||   :: ||:
Db        670 QELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVG 729

Qy        712 ARVQLSPPQLAYIVVTRSMNTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVSQQVR 767
                    |:|  ||| : :| |||        ||| |||::|  :| || |||::||| ||
Db        730 RGYN---PRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVR 786

Qy        768 QGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTN 827
               |:|||| |||:| | |  |: :|:|||||:||:|||| |  |||| |||| ||| ||| :
Db        787 SGSVSPTHYNVIYDSSGLKPDHIQRLTYKLCHVYYNWPGVIRVPAPCQYAHKLAFLVGQS 846

Qy        828 LHSIPQNALEKKFYYL 843
              :|  |  :|  : |||
Db        847 IHREPNLSLSNRLYYL 862


RESULT 9
Q8UVX0
ID   Q8UVX0      PRELIMINARY;      PRT;   858 AA.
AC   Q8UVX0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Piwi protein.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Weeraratne S.D., Gong Z., Tan C.-H.;
RT   "Cloning and characterization of zebrafish homolog of piwi, essential
RT   for germ-line stem cell self-renewal.";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF336369; AAL57170.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   858 AA;  97451 MW;  6A12F2E511465777 CRC64;

  Query Match           30.3%;  Score 1332;  DB 13;  Length 858;
  Best Local Similarity  33.5%;  Pred. No. 2.5e-87;
  Matches  299;  Conservative  189;  Mismatches  314;  Indels   90;  Gaps   23;

Qy          6 GRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREG- 64
              || | |         ||| |        ||     | | | :  |:|      :  ||
Db          3 GRARAR---------SRGRG--------RGQEPAAPGAQPPV--SQEAAKPVVSTPSEGQ 43

Qy         65 --GPPERKP------------------------WGDQYDY----LNTRP-VELV--SKK 90
                |   :||                        | : |:    ::||  :| |  ||
Db         44 LVGRGRQKPAPGAMSEEAMLQISAGFQQVKIGERGGRRRDFHDSGVHTRQLMEHVKESKT 103

Qy         91 GTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFVPTIENPRVRMGVLSNHANLLGSGYLFDG 150
              |  |  : |: || |  ::| | :  |||:: | :|: |:|  :|  |   ||  : |||
Db        104 GVSGTAIELRANFMRLLSRPMWALYQYHVDYKPPMESRRLRSALLFQHEETLGKAHTFDG 163

Qy        151 LQLFTTRKFEQEITVLSGKSKLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKGLNL 210
                ||   |     |||  :::   : :|::      :  : |  ||  |:: || :: ||:
Db        164 AILFLPNKLRNAETVLCSETRNGEKVEITVTLTNELPPSSPVCLQFYNILFRRILRILNM 223

Qy        211 ELVGRNLFDPRAKIEIREFKMELWPGYETSIRQHEKDILLGTEITHKVMRTETIYDIM-- 268
              : :||: ::|     | : :: :|||: |:| |:|  |:| ::::|||:|:||: | |
Db        224 QQIGRHYYNPDDPFNIPQHRLTIWPGFMTTILQYESSIMLCSDVSHKVLRSETVLDFMYS 283

Qy        269 --RRCSHNPARHQDEVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTFSCKGRDISFVE 326
                ::|     |  :     :: ||:|| |||:||||:|: :  || :||     :|||
Db        284 LRQQCGDQ--RFPEACTKELVGLIILTKYNNKTYRIDDIAWDHTPNNTFKKGDTEISFKN 341

Qy        327 YYLTKYNIRIRDHNQPLLISKNRDKALKTNASE---LVVLIPELCRVTGLNAEMRSNFQL 383
              |: ::| : | | || ||:|    | |  :         :|:|| | :|||  :||::| :
Db        342 YFKSQYGLDITDGNQVLLVS--HVKRLGPSGRPPPGPAMLVPEFCYLTGLTDKMRADFNI 399
```

```
Qy        384 MRAMSSYTRMNPKQRTDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNI 443
Db        400 MKDLASHTRLSPEQREGRINRLISNINRNGDVQNELTTWGLSFENKLLSLNGRVLPSERI 459

Qy        444 VFHNGKVPAGE----NADWQRHFRDQRMLTTPSDGLDRWAVIAPQRNSHELRTLLDSLYR 499
Db        460 I-QGGR--AFEYNPWTADWSKEMRGLPLISCMS--LDNWLMFYTRRNADVAQSLLQTLNK 514

Qy        500 AASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCVRSDPKLILCLVPNDNAERYSSIKKRGY 559
Db        515 VSGPMGIRMQR-AVMIEYEDRQESLLRALQQNVARETQMVVVILPTNRKDKYDCVKKYLC 573

Qy        560 VDRAVPTQVVTLKT-TKNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIAKS 618
Db        574 VDCPTPSQCVVSRTISKPQALMTVATKIALQMNCKMGGELWSVEIPLRQLMIVGIDCYHD 633

Qy        619 TRDRKRAYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRKLPS 678
Db        634 TAAGKRSIGAMVASLN-QGMSRWFSKCVLQNRGQEIIDALKGSLQGALKAYLKYNNSLPS 692

Qy        679 RIVFYRDGVSSGSLKQLFEFEVKDIIEKLKT---EYARVQLSPPQLAYIVVTRSMNTRFF 735
Db        693 RIIVYRDGVGDGMLQSVVDYEVPQIMQSIKTMGQDY------EPKLSVVVVKKRISSRFF 746

Qy        736 --LNGQ--NPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQ 791
Db        747 ARIDGKIANPPPGTVIDTEVTRPEWYDFFIVSQAVRFGCVAPTHYNVVFDNSGLKPDHMQ 806

Qy        792 KLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKFYYL 843
Db        807 RLTYKLCHMYYNWQGIVRVPAPCQYAHKLAFLVGQSIHKEPNMNLDDFLYYL 858


RESULT 10
Q9GPA7
ID   Q9GPA7      PRELIMINARY;      PRT;   808 AA.
AC   Q9GPA7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Seawi (Fragment).
OS   Strongylocentrotus purpuratus (Purple sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Strongylocentrotus.
OX   NCBI_TaxID=7668;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rodriguez A.J., Bonder E.M.;
RT   "Seawi - Cloning and Molecular Characterization of a Sea Urchin
RT   Homolog of Piwi.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY014900; AAG42534.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
FT   NON_TER     808    808
SQ   SEQUENCE   808 AA;  91366 MW;  19528F4B9D10474C CRC64;

  Query Match             30.2%;  Score 1326.5;  DB 5;  Length 808;
  Best Local Similarity   35.5%;  Pred. No. 5.8e-87;
  Matches  288;  Conservative  169;  Mismatches  298;  Indels   57;  Gaps   16;

Qy          6 GRGRRRPXNEDDSSTSRGSGDGPRVKVFRG----------SSSGDPRADPRIEASRERR 54
Db         30 GRGRSR-----GGTTQPGPGAVPKQEP-RGELTSQIAKLAIGASGEGRVGGAVQDRRRRR 83

Qy         55 -ALEEAPRREGGPPERKPWGDQYDYLNTRPVELVSKKGTDGVPVMLQTNFFRXKTKPEWR 113
Db         84 DEIELEP------------------TTRPENFV-KQAIAGDKIALIANGFKLKTKPDWQ 123

Qy        114 IVHYHVEFVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLD 173
```

```
Db        124 LYQYRVDFEPEILNPRARFALLKGHSALLGKGLTLDMDTMYSLFKLAEKVTNLSAERKDG 183

Qy        174 IEYKISIKFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMEL 233
Db        184 SNVNICVTHVATLNPMAPNTLHLYNVLFRRCLKMINMEQVGRNYYDPTAAIDIKQHGLQL 243

Qy        234 WPGYETSIRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNP-ARHQDEVRVNVLDLIVL 292
Db        244 WPGFVTSILQYEYDVMLLSDISHKVLRTQTVWEVMNDLFNKARGRFKEEITKLMIGQIVL 303

Qy        293 TDYNNRTYRINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKA 352
Db        304 TKYNNKTYRVDDIDFDTTPADTFETRSGPVSYVDYFKKSYERVIHDVNQPMLVSRPKKRE 363

Qy        353 LKTNASELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNT 412
Db        364 EKKGVGP-AYLPPELCFLTGLSDDMRADFNVMKDMGQHTRVGPQDRCRTLSGFIKKLSSN 422

Qy        413 PESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKV----PAGENADWQRHFRDQRML 468
Db        423 EEVKTYLDSWGMEFDKEQVKLTGRVLPPEKL-FQRGKQFSYNPS--NADWSRDTRGNAL- 478

Qy        469 TTPSDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAM 528
Db        479 -TDAKILNNWKIFYTRRDANRGQDFIKSLVRVANPMGMNVRGPEIVELPDDRTETYTRSL 537

Qy        529 DDCVRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKT-TKNRSLMSIATKIA 587
Db        538 QAQIAQDTQIVVVILPTNRKDRYDAIKKTCVVTHPCPSQVIVSRTLSKQQMLMSVATKIA 597

Qy        588 IQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTE 647
Db        598 MQMNCKMGGDLWRVEIPLSNLMIIGIDSYHDSLTKGRSVLGFVASMNKSQTSFFSSCAFQ 657

Qy        648 CSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKL 707
Db        658 HAQGEFGAN-LSTLMNNALKRYYQINEKFPERIIIFRDGVGDSQVNLVVDYELKQIKDTL 716

Qy        708 KTEYARVQLSPPQLAYIVVTRSMNTRFFLNGQ----NPPPGTIVDDVITLPERYDFYLVS 763
Db        717 DKVYP--QGTVHKLAVVVVKKRINNRFFANLRGGLSNPPPGTVIDDVVTKPHLYDYFIIS 774

Qy        764 QQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTY 795
Db        775 QSVRQGSVSPTSYNVVYDTTGLKPDHMQRLTY 806


RESULT 11
Q9JMB6
ID   Q9JMB6      PRELIMINARY;      PRT;   971 AA.
AC   Q9JMB6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   MILI (Miwi like).
GN   PIWIL2 OR PIWIL1L OR MILI.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Miyagawa S.K., Kimura T., Nakano T.;
RT   "Molecular Cloning and characterization of Piwi family genes.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB032605; BAA93706.1; -.
DR   MGD; MGI:1930036; Piwil2.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   971 AA;  109488 MW;  01E143C6513310FB CRC64;
```

Query Match 26.9%; Score 1182.5; DB 11; Length 971;
Best Local Similarity 32.2%; Pred. No. 1.9e-76;
Matches 277; Conservative 173; Mismatches 356; Indels 55; Gaps 18;

```
Qy          6 GRGRRR----PXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPR 61
              |||       |     ||  ||     |       | :: ||   ||:    :  ||
Db        143 GRGSSEVSLLPLGRAASSIGRGMDKPPSA---FGLTARDP---PRLP---QPPALSPTSL 193

Qy         62 REGGPP-----ERKPWGDQYDYLNTRPVELVSKKGTDGVPVMLQTNFFRXKTKPEWRIVH 116
              ||     |||              ||: |:|: | |  |  |  : :   |  :
Db        194 HSADPPPVLTMERK-----------EKELLVKQGSKGTPQSLGLNLIKIQCHNE-AVYQ 240

Qy        117 YHVEFVPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEY 176
              ||| | |::|   :| |:| :| :: |:   |||  |:   | :| : : | :    | |
Db        241 YHVTFSPSVECKSMRFGMLKDHQSVTGNVTAFDGSILYLPVKLQQVVELKSQRKTDDAEI 300

Qy        177 KISIKFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPG 236
              | |:   :      :   |:: || || |:::|||| :|| : : ::: ::::|||
Db        301 SIKIQLTKILEPCSDLCIPFYNVVFRRVMKLLDMKLVGRNFYDPTSAMVLQQHRLQIWPG 360

Qy        237 YETSIRQHEKDILLGTEITHKVMRTETIYDIMRRC-SHNPARHQDEVRVNVLDLIVLTDY 295
              |  |||: :  : |  :::|||:| ::: |:|      |   |||    ::  ||:| |
Db        361 YAASIRRTDGGLFLLADVSHKVIRNDSVLDVMHAIYQQNKEHFQDECSKLLVGSIVITRY 420

Qy        296 NNRTYRINDVDFGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISKNRDKALK 354
              |||||||:|||: :||| :|    |::|:|:||| | | ::: :|||||  :|    :
Db        421 NNRTYRIDDVDWNKTPKDSFVMSDGKEITFLEYYSKNYGITVKEDDQPLLI--HRPSERQ 478

Qy        355 TNASEL----VVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQ 410
              |   |    ::|:|||  :|||:  :|: :|: |: ::    ::|||    |     |:
Db        479 NNHGMLLKGEILLLPELSFMTGIPEKMKKDFRAMKDLTQQINLSPKQHHGALECLLQRIS 538

Qy        411 NTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHF-RDQRMLT 469
                :   |  | : | |:| :::||::  :  |    |      |: :| :   ||  :||
Db        539 QNETASNELTRWGLSLHKDVHKIEGRLLPMERINLRNTSFVTSEDLNWVKEVTRDASILT 598

Qy        470 TPSDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMD 529
              |  :  ||:  |:|   : | |:: | : |  :|:||  |    : |||  ||:| :
Db        599 IP---MHFWALFYPKRAMDQARELVNMLEKIAGPIGMRISPPAWVELKDDRIETYIRTIQ 655

Qy        530 DCVRSDPK--LILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKI 586
              :  : |  :::|::     : | :|||   |   ||:||: ::|  :   | |:| ||
Db        656 SLLGVEGKIQMVVCIIMGTRDDLYGAIKKLCCVQSPVPSQVINVRTIGQPTRLRSVAQKI 715

Qy        587 AIQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVT 646
              :|:|||||   | :::|| || || |:       |:   :|||::|     |   |
Db        716 LLQMNCKLGGELWGVDIPLKQLMVIGMDVYHDPSRGMRSVVGFVASINLTLTKWYSRVVF 775

Qy        647 ECSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEK 706
              :   ::: ::|   :  :|::|   :  || :|| |||||| | || :  :|    | :
Db        776 QMPHQEIV-DSLKLCLVGSLKKYYEVNHCLPEKIVVYRDGVSDGQLKTVANYE----IPQ 830

Qy        707 LKTEYARVQLSPPQLAYIVVTRSMNTRFFLNGQN----PPPGTIVDDVITLPERYDFYLV 762
              |:  :       |::   || : ::|  :|   :     | |||:||  ||  |  ||||:
Db        831 LQKCFEAFDNYHPKMVVFVVQKKISTNLYLAAPDHFVTPSPGTVVDHTITSCEWVDFYLL 890

Qy        763 SQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLAT 822
              :  ||||   || |  : ::   ||: ||:||:|:||:|:|| || |||| |:|| |||
Db        891 AHHVRQGCGIPTHYICVLNTANLSPDHMQRLTFKLCHMYWNWPGTIRVPAPCKYAHKLAF 950

Qy        823 LVGTNLHSIPQNALEKKFYYL 843
              | |  ||  |   |    ::|
Db        951 LSGQILHHEPAIQLCGNLFFL 971
```

RESULT 12
Q8TC59
ID Q8TC59 PRELIMINARY; PRT; 973 AA.
AC Q8TC59;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to piwi like homolog 1 (Drosophila)-like.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC025995; AAH25995.1; -.
SQ SEQUENCE 973 AA; 109848 MW; C44398136B144CA0 CRC64;

Query Match 26.5%; Score 1165; DB 4; Length 973;
Best Local Similarity 31.4%; Pred. No. 3.5e-75;
Matches 269; Conservative 176; Mismatches 349; Indels 62; Gaps 17;

```
Qy         22 RGSGDGPRVKVFRGSS------------SGDPRADPRIEASRERRALEEAPRREGGPP- 67
              ||| |   : : | :                |   |  |:: :     || ::|     |
Db        146 RGSSDASLLPLGRAAGGISREVDKPPCTFSTPSRGPPQLSSP---PALPQSPLHSPDRPL 202

Qy         68 ----ERKPWGDQYDYLNTRPVELVSKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFVP 123
                 | |                ||: |:|: | |  |  |  : :   |  : ||| | |
Db        203 VLTVEHK-----------EKELIVKQGSKGTPQSLGLNLVKIQCHNE-AVYQYHVTFSP 249

Qy        124 TIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIKFV 183
               :|   :| |:| :|  : |:   |||  |:   | :| : : | :     | | |:
Db        250 NVECKSMRFGMLKDHQAVTGNVTAFDGSILYLPVKLQQVLELKSQRKTDSAEISIKIQMT 309

Qy        184 GFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSIRQ 243
               :       :   |:: || || |:::|||| :|| : : ::: ::::|||  |||:
Db        310 KILEPCSDLCIPFYNVVFRRVMKLLDMKLVGRNFYDPTSAMVLQQHRLQIWPGYAASIRR 369

Qy        244 HEKDILLGTEITHKVMRTETIYDIMRRC-SHNPARHQDEVRVNVLDLIVLTDYNNRTYRI 302
               :  : |  :::|||:| : : |:|      |   |||    ::  ||:| ||||||||
Db        370 TDGGLFLLADVSHKVIRNDCVLDVMHAIYQQNKEHFQDECTKLLVGNIVITRYNNRTYRI 429

Qy        303 NDVDFGQTPKSTFSCK-GRDISFVEYYLTKYNIRIRDHNQPLLISK------NRDKALKT 355
              :|||: :||| :|:   |::|:|:|||   | | ::: :|||||  :       |   ||
Db        430 DDVDWNKTPKDSFTMSDGKEITFLEYYSKNYGITVKEEDQPLLIHRPSERQDNHGMLLKG 489

Qy        356 NASELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRAFNHRLQNTPES 415
              ::|:||| :||: :|: :|: |: ::    ::|||    |     |:     :
Db        490 E----ILLLPELSFMTGIPEKMKKDFRAMKDLAQQINLSPKQHHSALECLLQRIAKNEAA 545

Qy        416 VKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQRHF-RDQRMLTTPSDG 474
              |  | : | |:| :::||::  : |    |      :  :| :   ||  :|| |
Db        546 TNELMRWGLRLQKDVHKIEGRVLPMERINLKNTSFITSQELNWVKEVTRDPSILTIP--- 602

Qy        475 LDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCVRS 534
              :  ||:  |:|   : | |:: | : |  :|:|:  |    : |||  |||| :   : :
Db        603 MHFWALFYPKRAMDQARELVNMLEKIAGPIGMRMSPPAWVELKDDRIETYVRTIQSTLGA 662

Qy        535 DPK--LILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQLN 591
              : |  :::|::      : | :|||   |   ||:||| ::|  :   | |:| || :|:|
Db        663 EGKIQMVVCIIMGPRDDLYGAIKKLCCVQSPVPSQVVNVRTIGQPTRLRSVAQKILLQIN 722

Qy        592 CKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAF 651
              ||||   | :::||  || || |:       |:    :|||::|     |   | :
Db        723 CKLGGELWGVDIPLKQLMVIGMDVYHDPSRGMRSVVGFVASINLTLTKWYSRVVFQMPHQ 782

Qy        652 DVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEY 711
              ::: ::|   :  :|:::   :  || :|| |||||| | || :  :|:   ::|    :
Db        783 EIV-DSLKLCLVGSLKKFYEVNHCLPEKIVVYRDGVSDGQLKTVANYEIPQ-LQKCFEAF 840

Qy        712 ARVQLSPPQLAYIVVTRSMNTRFFL----NGQNPPPGTIVDDVITLPERYDFYLVSQQVR 767
              |   |::   || : ::|  :|    |   | |||:||  ||  |  ||||:: ||
Db        841 ENYQ---PKMVVFVVQKKISTNLYLAAPQNFVTPTPGTVVDHTITSCEWVDFYLLAHHVR 897

Qy        768 QGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTN 827
              ||   || |  : ::   ||: ||:||:|:||:|:|| || |||| |:|| ||| | |
Db        898 QGCGIPTHYVCVLNTANLSPDHMQRLTFKLCHMYWNWPGTIRVPAPCKYAHKLAFLSGHI 957
```

```
Qy        828 LHSIPQNALEKKFYYL 843
              ||  |   | :  ::|
Db        958 LHHEPAIQLCENLFFL 973


RESULT 13
P90786
ID   P90786      PRELIMINARY;      PRT;   824 AA.
AC   P90786;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   D2030.6 protein.
GN   D2030.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wilkinson J.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z73906; CAA98113.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   824 AA;  93844 MW;  970263CF6FEDBCEE CRC64;

  Query Match             24.8%;  Score 1087.5;  DB 5;  Length 824;
  Best Local Similarity   29.9%;  Pred. No. 1e-69;
  Matches  267;  Conservative  167;  Mismatches  341;  Indels  117;  Gaps   22;

Qy          1 MADDQGRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAP 60
              ||   ||||            ||||          ::||
Db          1 MASGSGRGR-----------GRGSG---------SNNSG---------------------- 19

Qy         61 RREGGPPERKPWGDQYDYLNTRPVELV------SKKGTDGVPVMLQTNFFRXK-TKPEWR 113
                        |    || |   :|       || |: | |     ||   : |: ::
Db         20 ------------GKDQKYLGTIQPDLFIRQQGQSKTGSSGQPQKCFANFIPIEMTQSDYS 67

Qy        114 IVHYHVEFVPTIENPRVRMGVL--SNHANLLGSGYLFDGLQLFTTRKFEQ----EITVLS 167
              |  ||||| ||:::   |  :|  :|  : :|  ::|||: |:   ::||    |:
Db         68 IYQYHVEFEPTVDSKANREKMLRDNNVTDEIGHHFVFDGMILYLKEEWEQNQMIEVQHPI 127

Qy        168 GKSKLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLF---DPRAKI 224
              :| : |:| : :|:      :|: : : | |:|||   | |  :||| |   | ||
Db        128 DRSLICIRFKQTNRFL----VDDPQTINIFNTIIRRSFDALQLTQLGRNYFNWGDSRA-- 181

Qy        225 EIREFKMELWPGYETSIRQHEKDILLGTEITHKVMRTETIYDI----MRRCSHNPARHQD 280
              : :: | : |||||:|| :|:: :|  |   |::| |::| :    :| | :|| | |:
Db        182 -VPDYNMSILPGYETAIRMYEENFMLCVENRFKMVREESMYILFHKELRSCQNNPQRVQE 240

Qy        281 EVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHN 340
              ::        ::| |||: :|   :|:  :| | |   |: |:  ||:  :| | |   :
Db        241 KMNEMYGGTTIITRYNNKLHRYTRLDYSISPLSEFVKDGQSITLKEYFKNQYGIEITVDD 300

Qy        341 QPLLISKNRDKALKTNASELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTD 400
              ||::||: : |       ::  ::|||||  |||  ||| :|::|: :: :|||:|:||
Db        301 QPIIISEGKPKQ-PGEPPQVSYIVPELCFPTGLTDEMRKDFKMMKEIAKHTRMSPQQRLV 359

Qy        401 RLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGENADWQR 460
              |     |    : ::  : | : | :::  || |::  : :     |   |: |:| |
Db        360 ESRKLIVDLSKNEKVMECFKYWGISLGQDLANVQARVLKSEPL--QGKKTYEGKQAEWAR 417

Qy        461 HFRDQRMLTTPSDGLDRWAVIAPQRNSHEL--RTLLDSLYRAASGMGLRIRSPQEFIIYD 518
                                                  ::  :       : | || |   :  | :  ::   |    : :::  |    |
Db        418 GVKECGIYR--GSNMTNWIVIGPGSGNSGLLSQKFIEEARRLGKILQVQLGEPMCVPIRG 475

Qy        519 DRTGTYVRAMDDCVR----SDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT 574
                    |:  :   ::      |  ::: :: :||  || |:||   |:  :| | | |:|
Db        476 ISPNDYLEGVKGAIKQVDGEDIHMLVVMLADDNKTRYDSLKKFLCVECPIPNQCVNLRTL 535

Qy        575 --------KNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAY 626
                     :|::| ||  || :|: || |   | : :||   | :|:|:  :  : :
Db        536 AGKSKDGGENKNLGSIVLKIVLQMICKTGGALWKVNIPLKNTMIVGYDLYHDSTLKGKTV 595

Qy        627 GALIA--SMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEH-RKLPSRIVFY 683
              || ::  | |  |   ::|          | | |   : |||:||   : : ||||:: |
Db        596 GACVSTTSNDFTQ---FYSQTRPHENPTQLGNNLTHFVRKALKQYYDSNDQTLPSRLILY 652

Qy        684 RDGVSSGSLKQLFEFEVK------DIIEKLKTEYARVQLSPPQLAYIVVTRSMNTRFFLN 737
              |||   | :  :  |||      | :      | :       :||:|:||: :| |
Db        653 RDGAGDGQIPYIKNTEVKLVRDACDAVTDKAAELSNKVQEKIKLAFIIVTKRVNMRILKQ 712

Qy        738 GQ------NPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQ 791
              |       || |||:||  :| ||| ||||| | | ||||:| |||:::      |:| |
Db        713 GSSLDNAINPQPGTVVDTTVTRPERMDFYLVPQFVNQGTVTPVSYNIIHDDTDLGPDKHQ 772

Qy        792 KLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKFYYL 843
              :| :|:|||||| || |||| |||| ||| |   :||      |  | ::|
Db        773 QLAFKLCHLYYNWQGTVRVPAPCQYAHKLAFLTAQSLHDDANGCLRDKLFFL 824


RESULT 14
Q17567
ID   Q17567      PRELIMINARY;      PRT;   722 AA.
AC   Q17567;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   C01G5.2 protein.
GN   C01G5.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Bradshaw H., Stellyes L.;
RT   "The sequence of C. elegans cosmid C01G5.";
RL   Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U50068; AAB37734.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   722 AA;  82651 MW;  35BA0117952AD61F CRC64;

  Query Match             21.7%;  Score 953.5;  DB 5;  Length 722;
  Best Local Similarity   29.8%;  Pred. No. 3.9e-60;
  Matches  230;  Conservative  144;  Mismatches  306;  Indels   91;  Gaps   19;

Qy        111 EWRIVHYHVEFVPTIENPRVRMGVLSNHANL--LGSGYLFDGLQLFTTRKFEQ----EIT 164
              :: |  |||||  ||:::   |  :|   :    :|  ::|||: |:   :::|    |:
```

```
Db          5 DYSIYQYHVEFEPTVDSKATRENMLRQPSVTVEIGKHFVFDGMILYLKEEWDQNQMIEVQ 64

Qy        165 VLSGKSKLDIEYKISIKFVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKI 224

Db         65 HPNDNSLICIRFKKTNRFL----VDDPQTINIFNTIIRRSFDAMKLTQIGRNYFD----- 115

Qy        225 EIREFKMELWPGYETSIRQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARHQDEVRV 284

Db        116 ---------WD-------------------NSRALRKE-----LRSCQNNRQRVQEKMNE 142

Qy        285 NVLDLIVLTDYNNRTYRINDVDFGQTPKSTFSCKGRDISFVEYYLTKYNIRIRDHNQPLL 344

Db        143 VYGGSTIITRYNNKLHRFTRLDNEITPLSKFQKDGEQIILKEYFKNQYDIDITDDEQFII 202

Qy        345 ISKNRDKALKTNASELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQRTDRLRA 404

Db        203 ISEGKPKQ-PGEPPQVNYIVPELCFPTGLTDEMRKDFKMMKEIAKHTRMSPQQRLDETRK 261

Qy        405 FNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVP-AGENADWQRHFR 463

Db        262 LITKLSQNQTMMECFQYWGISLGQDLANVQARVLKSEPL---QGKRQYEGKQAEWARGVK 318

Qy        464 DQRMLTTPSDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMG----LRIRSPQEFIIYDD 519

Db        319 ECGIYR--GSNMTNWIVIGPGSGNSGL--LAQKFIAEARNLGRTLQVQLGEPMCVKINGI 374

Qy        520 RTGTYVRAMDDCVRS----DPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT- 574

Db        375 SPNDYLEGLKAAIKSVDGEEIHMLVVMLADDNKTRYDSLKKYLCVECPIPNQCVNLRTLA 434

Qy        575 -------KNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYG 627

Db        435 GKSKDGGENKNLGSIVLKIVLQMICKTGGALWKVNIPLKSTMIVGYDLYHDSTLKGKTVG 494

Qy        628 ALIA--SMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRK-LPSRIVFYR 684

Db        495 ACVSTTSNDFTQ---FYSQTRPHENPTQLGNNLTHFVRKSLKQYYDNNDKTLPSRLILYR 551

Qy        685 DGVSSGSLKQLFEFEVK------DIIEKLKTEYARVQLSPPQLAYIVVTRSMNTRFFLNG 738

Db        552 DGAGDGQIPYIKNTEVKLVRDACDAVTDKAAELSNKVQEKIKLAFIIVTKRVNMRILKQG 611

Qy        739 Q------NPPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQK 792

Db        612 SSSKSAINPQPGTVVDTTVTRPERMDFYLVPQFVNQGTVTPVSYNIIHDDTGLGPDKHQQ 671

Qy        793 LTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKFYYL 843

Db        672 LAFKLCHLYYNWQGTVRVPAPCQYAHKLAFLTAQSLHDDANGYLRDKLFFL 722


RESULT 15
Q99MV6
ID   Q99MV6      PRELIMINARY;      PRT;   580 AA.
AC   Q99MV6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   PIWIL2.
GN   PIWIL1L OR PIWIL2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE=21175748; PubMed=11279525;
RA   Wang P.J., McCarrey J.R., Yang F., Page D.C.;
RT   "An abundance of X-linked genes expressed in spermatogonia.";
RL   Nat. Genet. 27:422-426(2001).
DR   EMBL; AF285586; AAK31965.1; -.
DR   MGD; MGI:1930036; Piwil2.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   580 AA;  66549 MW;  A4EA9DE8626FA136 CRC64;

  Query Match           19.9%;  Score 875.5;  DB 11;  Length 580;
  Best Local Similarity   33.8%;  Pred. No. 1.2e-54;
  Matches  197;  Conservative  120;  Mismatches  243;  Indels   23;  Gaps   10;

Qy        274 NPARHQDEVRVNVLDLIVLTDYNNRTYRINDVDFGQTPKSTF-SCKGRDISFVEYYLTKY 332

Db          8 NKEHFQDECSKLLVGSIVITRYNNRTYRIDDVDWNKTPKDSFVMSDGKEITFLEYYSKNY 67

Qy        333 NIRIRDHNQPLLISKNRDKALKTNASEL----VVLIPELCRVTGLNAEMRSNFQLMRAMS 388

Db         68 GITVKEDDQPLLI--HRPSERQNNHGMLLKGEILLLPELSFMTGIPEKMKKDFRAMKDLT 125

Qy        389 SYTRMNPKQRTDRLRAFNHRLQNTPESVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNG 448

Db        126 QQINLSPKQHHGALECLLQRISQNETASNELTRWGLSLHKDVHKIEGRLLPMERINLRNT 185

Qy        449 KVPAGENADWQRHF-RDQRMLTTPSDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLR 507

Db        186 SFVTSEGLNWVKEVTRDASILTIP---MHFWALFYPKRAMDQARELVNMLEKIAGPIGMR 242

Qy        508 IRSPQEFIIYDDRTGTYVRAMDDCVRSDPK--LILCLVPNDNAERYSSIKKRGYVDRAVP 565

Db        243 TSPPAWVELKDDRIETYIRTIQSLLGVEGKIQMVVCIIMGTRDDLYGAIKKLCCVQSPVP 302

Qy        566 TQVVTLKTT-KNRSLMSIATKIAIQLNCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKR 624

Db        303 SQVINVRTIGQPTRLRSVAQKILLQMNCKLGGELWGVDIPLKQLMVIGMDVYHDPSRGMR 362

Qy        625 AYGALIASMDLQQNSTYFSTVTECSAFDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYR 684

Db        363 SVVGFVASINLTLTKWYSRVVFQMPHQEIV-DSLKLCLVGSLKKYYEVNHCLPEKIVVYR 421

Qy        685 DGVSSGSLKQLFEFEVKDIIEKLKTEYARVQLSPPQLAYIVVTRSMNTRFFLNGQN---- 740

Db        422 DGVSDGQLKTVANYE----IPQLQKCFEAFDNYHPKMVVFVVQKKISTNLYLAAPDHFVT 477

Qy        741 PPPGTIVDDVITLPERYDFYLVSQQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHL 800

Db        478 PSPGTVVDHTITSCEWVDFYLLAHHVRQGCGIPTHYICVLNTANLSPDHMQRLTFKLCHM 537

Qy        801 YYNWSGTTRVPAVCQYAKKLATLVGTNLHSIPQNALEKKFYYL 843

Db        538 YWNWPGTIRVPAPCKYAHKLAFLSGQILHHEPAIQLCGNLFFL 580


Search completed: July 22, 2003, 07:04:47
Job time : 36.8527 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2003, 18:02:32 ; Search time 10243.5 Seconds
(without alignments)
11546.204 Million cell updates/sec

Title: US-09-873-737A-3
Perfect score: 4064
Sequence: 1 agaagcctgctacaggggt.....aaaaaaaaaaaaaaaaaaa 4064

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4058	99.9	4064	10	AF438405	AF438405 Mus muscu
	2	3859.8	95.0	3910	10	AB032604	AB032604 Mus muscu
	3	1835	45.2	3362	9	AF104260	AF104260 Homo sapi
	4	1835	45.2	3399	9	BC028581	BC028581 Homo sapi
	5	1835	45.2	3421	9	AF387507	AF387507 Homo sapi
	6	1742.8	42.9	2579	9	AK093133	AK093133 Homo sapi
	7	1680.6	41.4	2328	9	AF264004	AF264004 Homo sapi
	8	1288.6	31.7	236516	2	AC111089	AC111089 Mus muscu
	9	1119.2	27.5	3273	5	AF336369	AF336369 Danio rer
	10	734.8	18.1	3006	9	BC031060	BC031060 Homo sapi
	11	734.8	18.1	3138	9	AK093475	AK093475 Homo sapi
	12	663.8	16.3	3057	3	AY014899	AY014899 Strongylo
	13	660.2	16.2	2456	9	AK096837	AK096837 Homo sapi
	14	589.2	14.5	2605	3	AY014900	AY014900 Strongylo
	15	579.8	14.3	60131	2	AC116715	AC116715 Mus muscu
	16	474.8	11.7	4913	10	AB032605	AB032605 Mus muscu
	17	437.8	10.8	3610	9	BC025995	BC025995 Homo sapi
	18	433.2	10.7	3620	9	AK056418	AK056418 Homo sapi
	19	407.2	10.0	1723	3	AY014901	AY014901 Strongylo
	20	399.4	9.8	3649	6	AX247684	AX247684 Sequence
	21	399.4	9.8	3659	10	AF285586	AF285586 Mus muscu
	22	345	8.5	2272	9	AK027497	AK027497 Homo sapi
	23	311.8	7.7	3047	3	AF104354	AF104354 Drosophil
	24	286	7.0	2761	3	AF145680	AF145680 Drosophil
	25	259.6	6.4	1733	9	AK001213	AK001213 Homo sapi
c	26	185.8	4.6	39365	2	AC100141	AC100141 Mus muscu
	27	175	4.3	7781	3	DMDNARPL9	X94613 D.melanogas
c	28	175	4.3	37980	3	AC005116	AC005116 Drosophil
c	29	175	4.3	78507	2	AC019704	AC019704 Drosophil
c	30	175	4.3	169210	3	AC092226	AC092226 Drosophil
c	31	175	4.3	176306	3	AC007186	AC007186 Drosophil
c	32	175	4.3	276576	3	AE003630	AE003630 Drosophil
	33	173.4	4.3	901	3	AF334408	AF334408 Drosophil
	34	168.4	4.1	61995	2	AC090565	AC090565 Homo sapi
c	35	168.4	4.1	193697	2	AC025837	AC025837 Homo sapi
	36	168.4	4.1	198104	2	AC127071	AC127071 Homo sapi
	37	158.8	3.9	9359	3	AF104355	AF104355 Drosophil
	38	157.6	3.9	548	6	AX387649	AX387649 Sequence
	39	153.8	3.8	104649	2	AC096050	AC096050 Rattus no
c	40	149.8	3.7	104649	2	AC096050	AC096050 Rattus no
c	41	140.4	3.5	60131	2	AC116715	AC116715 Mus muscu
c	42	128.6	3.2	3677	2	AC017885	AC017885 Drosophil
	43	128.6	3.2	3855	3	AE003107	AE003107 Drosophil
	44	126.4	3.1	1034	9	AK026865	AK026865 Homo sapi
	45	126.2	3.1	1054	17	AF130104	Af130104 Homo sapi

ALIGNMENTS

RESULT 1
AF438405
LOCUS AF438405 4064 bp mRNA linear ROD 13-NOV-2001
DEFINITION Mus musculus MIWI mRNA, complete cds.
ACCESSION AF438405
VERSION AF438405.1 GI:16905060
KEYWORDS .
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4064)
AUTHORS Deng,W. and Lin,H.
TITLE Miwi, a murine homolog of piwi, encodes a cytoplasmic protein
essential for spermatogenesis

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4064)
  AUTHORS   Deng,W. and Lin,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2001) Department of Cell Biology, Duke
            University, Room 412, Nanaline Duke Building, Durham, NC 27710, USA
FEATURES             Location/Qualifiers
     source          1. .4064
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /tissue_type="testis"
     CDS             191. .2779
                     /note="Piwi-like protein"
                     /codon_start=1
                     /product="MIWI"
                     /protein_id="AAL31014.1"
                     /db_xref="GI:16905061"
                     /translation="MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEG
                     DLVGRGRQRGMVVGATSKSQELQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVK
                     ESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGR
                     CHAFDGTILFLPKRLQHKVTEVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIIFR
                     RLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLR
                     SETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFK
                     KADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTG
                     LTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNL
                     LSFSGRILQSEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEA
                     ANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTEAYLRALQQKVTSDTQIVVCLLSSNRK
                     DKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMPL
                     KLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQA
                     ALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIV
                     VKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVI
                     YDSSGLKPDHIQRLTYKLCHVYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLS
                     NRLYYL"
BASE COUNT     1115 a    982 c   1077 g    890 t
ORIGIN

  Query Match             99.9%;  Score 4058;  DB 10;  Length 4064;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 4058;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 AGAAGCCTGCTACAGGGGGTGGCGTGAGGAGCCTCAAGGACTGCGGACCTAAGCGTGCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGAAGCCTGCTACAGGGGGTGGCGTGAGGAGCCTCAAGGACTGCGGACCTAAGCGTGCTA 60

Qy         61 AGCACGGGGCCAGGATAGGACTGGACGCTGCTGGACCACGACGATCAGGGAGTGACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGCACGGGGCCAGGATAGGACTGGACGCTGCTGGACCACGACGATCAGGGAGTGACCCCC 120

Qy        121 ACCAGCATCCAGCCTGCGGCAGTTGTAGGACTGGAAGGAGGCCAAGCTGAACACCTGAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCAGCATCCAGCCTGCGGCAGTTGTAGGACTGGAAGGAGGCCAAGCTGAACACCTGAGC 180

Qy        181 TGACTGGAAAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGACTGGAAAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGA 240

Qy        241 GACGGTGCAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GACGGTGCAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACC 300

Qy        301 TCAACAGTCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCAACAGTCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGT 360

Qy        361 CGGAGCCACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGAGCCACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACT 420

Qy        421 GGCAGAGAGAGGAGGGCGTCGCCGAGACTTCCATGACNTTGGTGTGAACACCAGACAGAA 480
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        421 GGCAGAGAGAGGAGGGCGTCGCCGAGACTTCCATGACCTTGGTGTGAACACCAGACAGAA 480

Qy        481 CCTTGACCATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCTTGACCATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCAC 540

Qy        541 CAACCACTTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAACCACTTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTA 600

Qy        601 CAATCCCCTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAATCCCCTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCT 660

Qy        661 CATTGGAAGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CATTGGAAGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCA 720

Qy        721 CAAGGTCACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCAC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CAAGGTCACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCAC 780

Qy        781 CCTGACCAACGAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTT 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        781 CCTGACCAACGAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCATCTT 840

Qy        841 CAGGAGGCTCTTGAAAATCATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGGAGGCTCTTGAAAATCATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAG 900

Qy        901 TGACCCGATTGATATTCCAAACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TGACCCGATTGATATTCCAAACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCAT 960

Qy        961 CCTTCAGTATGAGAACAACATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCTTCAGTATGAGAACAACATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAG 1020

Qy       1021 CGAGACTGTCCTAGACTTCATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGAGACTGTCCTAGACTTCATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCA 1080

Qy       1081 GGAGCAAGTGTCGAAGGAGCTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGAGCAAGTGTCGAAGGAGCTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGAC 1140

Qy       1141 CTACCGGGTGGATGACATTGACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CTACCGGGTGGATGACATTGACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGA 1200

Qy       1201 TGGCTCGGAGGTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TGGCTCGGAGGTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGA 1260

Qy       1261 CCTGAAGCAGCCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CCTGAAGCAGCCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCT 1320

Qy       1321 GCCTGGCCCAGCTATGNTCATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAAT 1380
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GCCTGGCCCAGCTATGCTCATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAAT 1380

Qy       1381 GCGCAATGATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GCGCAATGATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCA 1440

Qy       1441 GCGGCAGCGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GCGGCAGCGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAG 1500

Qy       1501 AGAGCTTCGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAAT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AGAGCTTCGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAAT 1560

Qy       1561 CTTACAATCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGC 1620
```

```
Db      1561 CTTACAATCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGC 1620
Qy      1621 AGACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTG 1680
Db      1621 AGACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTG 1680
Qy      1681 GCTGCTGATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTT 1740
Db      1681 GCTGCTGATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTT 1740
Qy      1741 CAAAGTGACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGA 1800
Db      1741 CAAAGTGACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGA 1800
Qy      1801 CAGAACAGAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGT 1860
Db      1801 CAGAACAGAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGT 1860
Qy      1861 TGTCTGTCTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTG 1920
Db      1861 TGTCTGTCTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTG 1920
Qy      1921 TACAGACTGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAAC 1980
Db      1921 TACAGACTGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAAC 1980
Qy      1981 AGTCATGGCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCT 2040
Db      1981 AGTCATGGCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCT 2040
Qy      2041 CTGGCGGGTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGA 2100
Db      2041 CTGGCGGGTGGACATGCCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGA 2100
Qy      2101 CACCACAGCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGAC 2160
Db      2101 CACCACAGCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGAC 2160
Qy      2161 CCGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAA 2220
Db      2161 CCGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAA 2220
Qy      2221 GGTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCG 2280
Db      2221 GGTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCG 2280
Qy      2281 TGTCATCGTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGA 2340
Db      2281 TGTCATCGTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGA 2340
Qy      2341 GGTCCCACAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGAC 2400
Db      2341 GGTCCCACAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGAC 2400
Qy      2401 TGTAATCGTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACT 2460
Db      2401 TGTAATCGTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACT 2460
Qy      2461 TCAGAACCCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGA 2520
Db      2461 TCAGAACCCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGA 2520
Qy      2521 CTTTTTCATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGT 2580
Db      2521 CTTTTTCATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGT 2580
Qy      2581 CATCTATGACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCTG 2640
Db      2581 CATCTATGACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGCTCTG 2640
Qy      2641 CCACGTGTACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACA 2700
```

```
Db      2641 CCACGTGTACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACA 2700
Qy      2701 CAAGCTGGCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAA 2760
Db      2701 CAAGCTGGCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAA 2760
Qy      2761 CCGCCTCTACTACCTCTAACCAGCAGCGAGGGCCGGCATGCTGGTTTCCTTCAGAGCTCT 2820
Db      2761 CCGCCTCTACTACCTCTAACCAGCAGCGAGGGCCGGCATGCTGGTTTCCTTCAGAGCTCT 2820
Qy      2821 GCCCAATCCTTCAGTTCTATTTTTCATAGGATGAGGTTCTAGCGGGATGTTGGGGGTAGG 2880
Db      2821 GCCCAATCCTTCAGTTCTATTTTTCATAGGATGAGGTTCTAGCGGGATGTTGGGGGTAGG 2880
Qy      2881 GGTTGGGGGCCGCTGTTTTGTTTTGTTTTGTTTTTAGTTGTTGGTGTGTAGCGGCTATAG 2940
Db      2881 GGTTGGGGGCCGCTGTTTTGTTTTGTTTTGTTTTTAGTTGTTGGTGTGTAGCGGCTATAG 2940
Qy      2941 AGTTAACTAAGGCTTATGTTTGCTTTGTTTCTCATGGTGTGTTGTGACAGTTTGCTTGGC 3000
Db      2941 AGTTAACTAAGGCTTATGTTTGCTTTGTTTCTCATGGTGTGTTGTGACAGTTTGCTTGGC 3000
Qy      3001 AGAGTGGTGGCAAGGGCCGTGGTACCACAGATAGACTTTCCTAAAGCAGCCGTGTTGGCA 3060
Db      3001 AGAGTGGTGGCAAGGGCCGTGGTACCACAGATAGACTTTCCTAAAGCAGCCGTGTTGGCA 3060
Qy      3061 GGTTCTTGGAGGTGCTCCTGGGAGCACCTCCTATTGTACGCTGCTCTCTCTCTCTCTCTC 3120
Db      3061 GGTTCTTGGAGGTGCTCCTGGGAGCACCTCCTATTGTACGCTGCTCTCTCTCTCTCTCTC 3120
Qy      3121 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGCATGTTTGAATTTAAAAGAGGA 3180
Db      3121 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGCATGTTTGAATTTAAAAGAGGA 3180
Qy      3181 AAGGACACAAGGCAAGAATGCCCAGAGCATCCTAGATGGGTGCTATATGGGACTCTGACA 3240
Db      3181 AAGGACACAAGGCAAGAATGCCCAGAGCATCCTAGATGGGTGCTATATGGGACTCTGACA 3240
Qy      3241 CCTCAGGTTCTTAAAGATAAAAAGGCCAGAGGTACCTTTCTGTAATTCTTTTAAGAGTTA 3300
Db      3241 CCTCAGGTTCTTAAAGATAAAAAGGCCAGAGGTACCTTTCTGTAATTCTTTTAAGAGTTA 3300
Qy      3301 GGTCATAAGGAATGAAAAGAGTTACTTTTCTCCCTCCATGAAAGTGTTTTTTAAGACTGA 3360
Db      3301 GGTCATAAGGAATGAAAAGAGTTACTTTTCTCCCTCCATGAAAGTGTTTTTTAAGACTGA 3360
Qy      3361 CAAGGCCGGTGGTGGAGTGAGTGTTGTTGTTTTGTTTCGCAATTTTCTTACCAAGTGAAA 3420
Db      3361 CAAGGCCGGTGGTGGAGTGAGTGTTGTTGTTTTGTTTCGCAATTTTCTTACCAAGTGAAA 3420
Qy      3421 TTACTGACCATAATTTTATTTAATAGTTTTTAGTTTTTGAAAGCCTGTGTTTTTCTTTGCT 3480
Db      3421 TTACTGACCATAATTTTATTTAATAGTTTTTAGTTTTTGAAAGCCTGTGTTTTTCTTTGCT 3480
Qy      3481 TTTAAACTACAGGCGACCACTCTGGGATAGGATTCCACTCCCAAAAGTAATGTGTTGGGT 3540
Db      3481 TTTAAACTACAGGCGACCACTCTGGGATAGGATTCCACTCCCAAAAGTAATGTGTTGGGT 3540
Qy      3541 GTGGTGGCACACGCCTGTGATCTCAGGACTCAGGAGGTGAAGCCAGGGGAACCCCAAGTT 3600
Db      3541 GTGGTGGCACACGCCTGTGATCTCAGGACTCAGGAGGTGAAGCCAGGGGAACCCCAAGTT 3600
Qy      3601 CCCGGTCAGCCTAGGCTACATAGTGAGACTTTGTCTTAACAAACAGAAAAAAAGAAAAAG 3660
Db      3601 CCCGGTCAGCCTAGGCTACATAGTGAGACTTTGTCTTAACAAACAGAAAAAAAGAAAAAG 3660
Qy      3661 AAAGAAAAAAAAAAAGAAAGAAAAAAAAAARAAAGAAAAATGGAGCAGTATGAGTTATGGGT 3720
Db      3661 AAAGAAAAAAAAAAAGAAAGAAAAAAAAAGAAAGAAAAATGGAGCAGTATGAGTTATGGGT 3720
Qy      3721 GCCAGTTTCTGCAGGAGACTTTGGGAGTAAGAGCTGCCTCTGTTCCTAAGAGGCTGCAAG 3780
Db      3721 GCCAGTTTCTGCAGGAGACTTTGGGAGTAAGAGCTGCCTCTGTTCCTAAGAGGCTGCAAG 3780
```

```
Qy       3781 GCTGCGGCTGGGGGGCCAGTTCAGGTGCCAGGTGAGGTACGGCCACATCTATTTGCCTGG 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 GCTGCGGCTGGGGGGCCAGTTCAGGTGCCAGGTGAGGTACGGCCACATCTATTTGCCTGG 3840

Qy       3841 CTCTTTGACAAAGAGAGCAGCTTAGAGTGGCTAGAGTGTGATGAGTTTTACACAATGGCG 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3841 CTCTTTGACAAAGAGAGCAGCTTAGAGTGGCTAGAGTGTGATGAGTTTTACACAATGGCG 3900

Qy       3901 ATGTGTGTGTGTGCCCCTGGCACAGCTCAATAAAGTTTCTGTCTCACTTTAAAAAAAAAA 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3901 ATGTGTGTGTGTGCCCCTGGCACAGCTCAATAAAGTTTCTGTCTCACTTTAAAAAAAAAA 3960

Qy       3961 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3961 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4020

Qy       4021 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4064
             ||||||||||||||||||||||||||||||||||||||||||||||
Db       4021 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4064
```

RESULT 2
AB032604
LOCUS AB032604 3910 bp mRNA linear ROD 04-OCT-2001
DEFINITION Mus musculus mRNA for MIWI, complete cds.
ACCESSION AB032604
VERSION AB032604.1 GI:7416110
KEYWORDS MIWI (piwi).
SOURCE Mus musculus tissue_lib:testis cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Kuramochi-Miyagawa,S., Kimura,T., Yomogida,K., Kuroiwa,A., Tadokoro,Y., Fujita,Y., Sato,M., Matsuda,Y. and Nakano,T.
TITLE Two mouse piwi-related genes: miwi and mili
JOURNAL Mech. Dev. 108 (1-2), 121-133 (2001)
MEDLINE 21463379
REFERENCE 2 (bases 1 to 3910)
AUTHORS Miyagawa,S.K. and Nakano,T.
TITLE Direct Submission
JOURNAL Submitted (21-SEP-1999) Satomi K Miyagawa, Research Institute for Microbial Diseases, Department of Molecular Cell Biology; Yamadaoka 3-1, Suita, Osaka 565-0871, Japan (E-mail:smiyagaw@biken.osaka-u.ac.jp, Tel:81-6-6879-8362, Fax:81-6-6879-8362)
FEATURES Location/Qualifiers
source 1. .3910
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="5F distal-G2 proximal"
/tissue_lib="testis"
gene 1. .3910
/gene="miwi"
CDS 156. .2744
/gene="miwi"
/note="piwi family"
/codon_start=1
/product="MIWI"
/protein_id="BAA93705.1"
/db_xref="GI:7416111"
/translation="MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEG
DLVGRGRQRGMVVGATSKSQELQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVK
ESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGR
CHAFDGTILFLPKRLQHKVTEVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIIFR
RLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLR
SETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFK
KADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTG
LTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNL
LSFSGRILQSEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEA
ANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTEAYLRALQQKVTSDTQIVVCLLSSNRK
DKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMPL
KLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQA
ALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIV
VKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVI
YDSSGLKPDHIQRLTYKLCHVYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLS
NRLYYL"
BASE COUNT 993 a 970 c 1063 g 884 t
ORIGIN

Query Match 95.0%; Score 3859.8; DB 10; Length 3910;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 3900; Conservative 1; Mismatches 7; Indels 9; Gaps 3;

```
Qy         36 AAGGACTGCGGACCTAAGCGTGCTAAGCACGGGGCCAGGATAGGACTGGACGCTGCTGGA 95
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGGACTGCGGACCTAAGCGTGCTAAGCACGGGGCCAGGATAGGACTGGACGCTGCTGGA 60

Qy         96 CCACGACGATCAGGGAGTGACCCCCACCAGCATCCAGCCTGCGGCAGTTGTAGGACTGGA 155
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCACGACGATCAGGGAGTGACCCCCACCAGCATCCAGCCTGCGGCAGTTGTAGGACTGGA 120

Qy        156 AGGAGGCCAAGCTGAACACCTGAGCTGACTGGAAAATGACTGGCCGAGCCCGAGCTCGGG 215
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGGAGGCCAAGCTGAACACCTGAGCTGACTGGAAAATGACTGGCCGAGCCCGAGCTCGGG 180

Qy        216 CCCGCGGCAGGGCACGAGGTCAGGAGACGGTGCAGCATGTTGGGGCTGCTGCGAGCCAGC 275
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCCGCGGCAGGGCACGAGGTCAGGAGACGGTGCAGCATGTTGGGGCTGCTGCGAGCCAGC 240

Qy        276 AACCTGGGTACATCCCACCGAGACCTCAACAGTCCCCCACAGAGGGGGACTTGGTTGGCC 335
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACCTGGGTACATCCCACCGAGACCTCAACAGTCCCCCACAGAGGGGGACTTGGTTGGCC 300

Qy        336 GAGGACGACAGAGGGGGATGGTAGTCGGAGCCACATCCAAGTCACAAGAACTGCAGATCT 395
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGGACGACAGAGGGGGATGGTAGTCGGAGCCACATCCAAGTCACAAGAACTGCAGATCT 360

Qy        396 CAGCTGGGTTTCAGGAGCTGTCACTGGCAGAGAGAGGAGGGCGTCGCCGAGACTTCCATG 455
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAGCTGGGTTTCAGGAGCTGTCACTGGCAGAGAGAGGAGGGCGTCGCCGAGACTTCCATG 420

Qy        456 ACNTTGGTGTGAACACCAGACAGAACCTTGACCATGTCAAAGAGTCAAAGACAGGCTCCT 515
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACCTTGGTGTGAACACCAGACAGAACCTTGACCATGTCAAAGAGTCAAAGACAGGCTCCT 480

Qy        516 CTGGCATCATTGTGAAGCTGAGCACCAACCACTTCCGGCTGACCTCGCGCCCACAGTGGG 575
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGGCATCATTGTGAAGCTGAGCACCAACCACTTCCGGCTGACCTCGCGCCCACAGTGGG 540

Qy        576 CCCTGTATCAGTACCACATCGACTACAATCCCCTGATGGAGGCCCGAAGGCTTCGCTCCG 635
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCCTGTATCAGTACCACATCGACTACAATCCCCTGATGGAGGCCCGAAGGCTTCGCTCCG 600

Qy        636 CACTGCTCTTCCAGCATGAAGACCTCATTGGAAGGTGTCATGCTTTCGATGGGACAATAT 695
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CACTGCTCTTCCAGCATGAAGACCTCATTGGAAGGTGTCATGCTTTCGATGGGACAATAT 660

Qy        696 TGTTTTTACCTAAGAGACTACAGCACAAGGTCACAGAAGTATTCAGTCAGACTCGGAATG 755
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TGTTTTTACCTAAGAGACTACAGCACAAGGTCACAGAAGTATTCAGTCAGACTCGGAATG 720

Qy        756 GGGAACACGTGAGGATCACCATCACCCTGACCAACGAGCTGCCGCCCACCTCGCCCACCT 815
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGGAACACGTGAGGATCACCATCACCCTGACCAACGAGCTGCCGCCCACCTCGCCCACCT 780

Qy        816 GCCTGCAGTTCTATAACATCNTCTTCAGGAGGCTCTTGAAAATCATGAATTTGCAACAAA 875
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCTGCAGTTCTATAACATCATCTTCAGGAGGCTCTTGAAAATCATGAATTTGCAACAAA 840

Qy        876 TTGGACGGAATTATTACAATCCAAGTGACCCGATTGATATTCCAAACCACAGGTTGGTGA 935
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTGGACGGAATTATTACAATCCAAGTGACCCGATTGATATTCCAAACCACAGGTTGGTGA 900

Qy        936 TCTGGCCCGGCTTCACCACCTCCATCCTTCAGTATGAGAACAACATCATGCTCTGCACAG 995
```

```
Db        901 TCTGGCCCGGCTTCACCACCTCCATCCTTCAGTATGAGAACAACATCATGCTCTGCACAG 960
Qy        996 ACGTCAGCCACAAGGTGCTCCGCAGCGAGACTGTCCTAGACTTCATGTTCAATCTATACC 1055
Db        961 ACGTCAGCCACAAGGTGCTCCGCAGCGAGACTGTCCTAGACTTCATGTTCAATCTATACC 1020
Qy       1056 AGCAGACAGAGGAGCACAAGTTCCAGGAGCAAGTGTCGAAGGAGCTCATAGGCCTCATCG 1115
Db       1021 AGCAGACAGAGGAGCACAAGTTCCAGGAGCAAGTGTCGAAGGAGCTCATAGGCCTCATCG 1080
Qy       1116 TTCTCACCAAGTACAATAACAAGACCTACCGGGTGGATGACATTGACTGGGACCAGAATC 1175
Db       1081 TTCTCACCAAGTACAATAACAAGACCTACCGGGTGGATGACATTGACTGGGACCAGAATC 1140
Qy       1176 CAAAGAGCACCTTCAAGAAGGCGGATGGCTCGGAGGTCAGCTTCCTGGAGTACTACAGGA 1235
Db       1141 CAAAGAGCACCTTCAAGAAGGCGGATGGCTCGGAGGTCAGCTTCCTGGAGTACTACAGGA 1200
Qy       1236 AGCAATACAACCAGGAGATCACGGACCTGAAGCAGCCGGTGCTGGTGAGCCAACCCAAGC 1295
Db       1201 AGCAATACAACCAGGAGATCACGGACCTGAAGCAGCCGGTGCTGGTGAGCCAACCCAAGC 1260
Qy       1296 GGAGGAGAGGCCCCGGCGGCACCCTGCCTGGCCCAGCTATGNTCATCCCTGAACTCTGCT 1355
Db       1261 GGAGGAGAGGCCCCGGCGGCACCCTGCCTGGCCCAGCTATGCTCATCCCTGAACTCTGCT 1320
Qy       1356 ATCTCACAGGCCTGACTGATAAAATGCGCAATGATTTCAATGTGATGAAGGACCTGGCAG 1415
Db       1321 ATCTCACAGGCCTGACTGATAAAATGCGCAATGATTTCAATGTGATGAAGGACCTGGCAG 1380
Qy       1416 TGCACACGCGGCTGACCCCTGAGCAGCGGCAGCGGGAGGTGGGCCGCCTCATCGACTACA 1475
Db       1381 TGCACACGCGGCTGACCCCTGAGCAGCGGCAGCGGGAGGTGGGCCGCCTCATCGACTACA 1440
Qy       1476 TCCACAAGGATGACAATGTGCAGAGAGAGCTTCGAGACTGGGGCCTGAGCTTCGACTCAA 1535
Db       1441 TCCACAAGGATGACAATGTGCAGAGAGAGCTTCGAGACTGGGGCCTGAGCTTCGACTCAA 1500
Qy       1536 ACTTGCTGTCCTTCTCTGGAAGAATCTTACAATCTGAGAAGATCCACCAGGGCGGAAAGA 1595
Db       1501 ACTTGCTGTCCTTCTCTGGAAGAATCTTACAATCTGAGAAGATCCACCAGGGCGGAAAGA 1560
Qy       1596 CGTTTGATTACAACCCACAATTTGCAGACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGA 1655
Db       1561 CGTTTGATTACAACCCACAATTTGCAGACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGA 1620
Qy       1656 TCAGCGTGAAGCCATTGGATAACTGGCTGCTGATCTATACCCGCAGGAATTATGAAGCAG 1715
Db       1621 TCAGCGTGAAGCCATTGGATAACTGGCTGCTGATCTATACCCGCAGGAATTATGAAGCAG 1680
Qy       1716 CCAACTCACTGATACAGAACCTGTTCAAAGTGACTCCAGCCATGGGCATCCAGATGAAAA 1775
Db       1681 CCAACTCACTGATACAGAACCTGTTCAAAGTGACTCCAGCCATGGGCATCCAGATGAAAA 1740
Qy       1776 AGGCAATCATGATCGAGGTGGATGACAGAACAGAAGCTTATCTGAGAGCCTTGCAGCAGA 1835
Db       1741 AGGCAATCATGATCGAGGTGGATGACAGAACAGAAGCTTATCTGAGAGCCTTGCAGCAGA 1800
Qy       1836 AGGTGACGTCAGACACTCAGATAGTTGTCTGTCTCTTGTCAAGTAATCGGAAGGACAAAT 1895
Db       1801 AGGTGACGTCAGACACTCAGATAGTTGTCTGTCTCTTGTCAAGTAATCGGAAGGACAAAT 1860
Qy       1896 ATGATGCCATCAAGAAGTACTTGTGTACAGACTGCCCCACCCCAAGTCAGTGTGTGGTGG 1955
Db       1861 ATGATGCCATCAAGAAGTACTTGTGTACAGACTGCCCCACCCCAAGTCAGTGTGTGGTGG 1920
Qy       1956 CCCGGACCCTGGGCAAGCAGCAAACAGTCATGGCCATTGCCACCAAGATCGCCCTGCAGA 2015
Db       1921 CCCGGACCCTGGGCAAGCAGCAAACAGTCATGGCCATTGCCACCAAGATCGCCCTGCAGA 1980
Qy       2016 TGAACTGCAAGATGGGAGGCGAGCTCTGGCGGGTGGACATGGCCCTGAAACTGGCAATGA 2075
Db       1981 TGAACTGCAAGATGGGAGGCGAGCTCTGGCGGGTGGACATGCCCCTGAAACTGGCAATGA 2040
Qy       2076 TCGTGGGCATCGACTGTTACCATGACACCACAGCTGGGCGGAGGTCCATCGCAGGATTCG 2135
Db       2041 TCGTGGGCATCGACTGTTACCATGACACCACAGCTGGGCGGAGGTCCATCGCAGGATTCG 2100
Qy       2136 TCGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCG 2195
Db       2101 TCGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCG 2160
Qy       2196 GGCAGGAGCTGGTGGATGGTCTCAAGGTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTG 2255
Db       2161 GGCAGGAGCTGGTGGATGGTCTCAAGGTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTG 2220
Qy       2256 GCTGCAATGAATACATGCCCAGCCGTGTCATCGTGTACCGAGACGGTGTGGGGGACGGGC 2315
Db       2221 GCTGCAATGAATACATGCCCAGCCGTGTCATCGTGTACCGAGACGGTGTGGGGGACGGGC 2280
Qy       2316 AGCTGAAGACCCTGGTCAATTATGAGGTCCCACAGTTCCTAGATTGCCTCAAGTCAGTCG 2375
Db       2281 AGCTGAAGACCCTGGTCAATTATGAGGTCCCACAGTTCCTAGATTGCCTCAAGTCAGTCG 2340
Qy       2376 GGAGAGGTTACAACCCAAGACTGACTGTAATCGTGGTGAAGAAGCGTGTCAATGCCAGGT 2435
Db       2341 GGAGAGGTTACAACCCAAGACTGACTGTAATCGTGGTGAAGAAGCGTGTCAATGCCAGGT 2400
Qy       2436 TTTTTGCTCAGTCTGGGGGAAGACTTCAGAACCCTCTTCCAGGGACAGTCATCGATGTGG 2495
Db       2401 TTTTTGCTCAGTCTGGGGGAAGACTTCAGAACCCTCTTCCAGGGACAGTCATCGATGTGG 2460
Qy       2496 AAGTCACCAGACCAGAGTGGTATGACTTTTTCATCGTGAGCCAGGCAGTGAGAAGCGGGA 2555
Db       2461 AAGTCACCAGACCAGAGTGGTATGACTTTTTCATCGTGAGCCAGGCAGTGAGAAGCGGGA 2520
Qy       2556 GTGTGTCCCCAACACACTACAATGTCATCTATGACAGCAGTGGCCTGAAGCCCGACCACA 2615
Db       2521 GTGTGTCCCCAACACACTACAATGTCATCTATGACAGCAGTGGCCTGAAGCCCGACCACA 2580
Qy       2616 TCCAGCGGCTGACATACAAGNTCTGCCACGTGTACTATAATTGGCCTGGAGTCATCCGAG 2675
Db       2581 TCCAGCGGCTGACATACAAGCTCTGCCACGTGTACTATAATTGGCCTGGAGTCATCCGAG 2640
Qy       2676 TCCCTGCACCTTGCCAGTATGCACACAAGCTGGCCTTCCTCGTGGGCCAGAGCATCCACA 2735
Db       2641 TCCCTGCACCTTGCCAGTATGCACACAAGCTGGCCTTCCTCGTGGGCCAGAGCATCCACA 2700
Qy       2736 GAGAGCCAAACCTCTCCCTGTCCAACCGCCTCTACTACCTCTAACCAGCAGCGAGGGCCG 2795
Db       2701 GAGAGCCAAACCTCTCCCTGTCCAACCGCCTCTACTACCTCTAACCAGCAGCGAGGGCCG 2760
Qy       2796 GCATGCTGGTTTCCTTCAGAGCTCTGCCCAATCCTTCAGTTCTATTTTTCATAGGATGAG 2855
Db       2761 GCATGCTGGTTTCCTTCAGAGCTCTG-CCAATCCTTCAGTTCTATTTTTCATAGGATGAG 2819
Qy       2856 GTTCTAGCGGGATGTTGGGGGTAGGGGTTGGGGGCCGCTGTTTTGTTTTGTTTTGTTTTT 2915
Db       2820 GTTCTAGCGGGATGTTGGGGGTAGGGGTTGGGGGCCGCTGTTTTGTTTTGTTTTGTTTTT 2879
Qy       2916 AGTTGTTGGTGTGTAGCGGCTATAGAGTTAACTAAGGCTTATGTTTGCTTTGTTTCTCAT 2975
Db       2880 AGTTGTTGGTGTGTAGCGGCTATAGAGTTAACTAAGGCTTATGTTTGCTTTGTTTCTCAT 2939
Qy       2976 GGTGTGTTGTGACAGTTTGCTTGGCAGAGTGGTGGCAAGGGCCGTGGTACCACAGATAGA 3035
Db       2940 GGTGTGTTGTGACAGTTTGCTTGGCAGAGTGGTGGCAAGGGCCGTGGTACCACAGATAGA 2999
Qy       3036 CTTTCCTAAAGCAGCCGTGTTGGCAGGTTCTTGGAGGTGCT--CCTGGGAGCACCTCCTA 3093
Db       3000 CTTTCCTAAAGCAGCCGTGTTGGCAGGTTCTTGGAGGTGCTTCCTTGGGAGCACCTCCTA 3059
Qy       3094 TTGTACGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3153
Db       3060 TTGTACGCTG------CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3113
```

```
Qy       3154 CTCTGCATGTTTGAATTTAAAAGAGGAAAGGACACAAGGCAAGAATGCCCAGAGCATCCT 3213
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3114 CTCTGCATGTTTGAATTTAAAAGAGGAAAGGACACAAGGCAAGAATGCCCAGAGCATCCT 3173

Qy       3214 AGATGGGTGCTATATGGGACTCTGACACCTCAGGTTCTTAAAGATAAAAAGGCCAGAGGT 3273
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3174 AGATGGGTGCTATATGGGACTCTGACACCTCAGGTTCTTAAAGATAAAAAGGCCAGAGGT 3233

Qy       3274 ACCTTTCTGTAATTCTTTTAAGAGTTAGGTCATAAGGAATGAAAAGAGTTACTTTTCTCC 3333
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3234 ACCTTTCTGTAATTCTTTTAAGAGTTAGGTCATAAGGAATGAAAAGAGTTACTTTTCTCC 3293

Qy       3334 CTCCATGAAAGTGTTTTTTAAGACTGACAAGGCCGGTGGTGGAGTGAGTGTTGTTGTTTT 3393
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3294 CTCCATGAAAGTGTTTTTTAAGACTGACAAGGCCGGTGGTGGAGTGAGTGTTGTTGTTTT 3353

Qy       3394 GTTTCGCAATTTTCTTACCAAGTGAAATTACTGACCATAATTTTATTTAATAGTTTTTAG 3453
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3354 GTTTCGCAATTTTCTTACCAAGTGAAATTACTGACCATAATTTTATTTAATAGTTTTTAG 3413

Qy       3454 TTTTTGAAAGCCTGTGTTTTCTTTGCTTTTAAACTACAGGCGACCACTCTGGGATAGGAT 3513
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3414 TTTTTGAAAGCCTGTGTTTTCTTTGCTTTTAAACTACAGGCGACCACTCTGGGATAGGAT 3473

Qy       3514 TCCACTCCCAAAAGTAATGTGTTGGGTGTGGTGGCACACGCCTGTGATCTCAGGACTCAG 3573
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3474 TCCACTCCCAAAAGTAATGTGTTGGGTGTGGTGGCACACGCCTGTGATCTCAGGACTCAG 3533

Qy       3574 GAGGTGAAGCCAGGGGAACCCCAAGTTCCCGGTCAGCCTAGGCTACATAGTGAGACTTTG 3633
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3534 GAGGTGAAGCCAGGGGAACCCCAAGTTCCCGGTCAGCCTAGGCTACATAGTGAGACTTTG 3593

Qy       3634 TCTTAACAAACAGAAAAAAAGAAAAAGAAAGAAAAAAAAAAGAAAGAAAAAAAAARAAAG 3693
             ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db       3594 TCTTAACAAACAGAAAAAAAGAAAAAGAAAGAAAAAAAAAAGAAAGAAAAAAAAAGAAAG 3653

Qy       3694 AAAAATGGAGCAGTATGAGTTATGGGTGCCAGTTTCTGCAGGAGACTTTGGGAGTAAGAG 3753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3654 AAAAATGGAGCAGTATGAGTTATGGGTGCCAGTTTCTGCAGGAGACTTTGGGAGTAAGAG 3713

Qy       3754 CTGCCTCTGTTCCTAAGAGGCTGCAAGGCTGCGGCTGGGGGGCCAGTTCAGGTGCCAGGT 3813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3714 CTGCCTCTGTTCCTAAGAGGCTGCAAGGCTGCGGCTGGGGGGCCAGTTCAGGTGCCAGGT 3773

Qy       3814 GAGGTACGGCCACATCTATTTGCCTGGCTCTTTGACAAAGAGAGCAGCTTAGAGTGGCTA 3873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3774 GAGGTACGGCCACATCTATTTGCCTGGCTCTTTGACAAAGAGAGCAGCTTAGAGTGGCTA 3833

Qy       3874 GAGTGTGATGAGTTTTACACAATGGCGATGTGTGTGTGTGCCCCTGGCACAGCTCAATAA 3933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3834 GAGTGTGATGAGTTTTACACAATGGCGATGTGTGTGTGTGCCCCTGGCACAGCTCAATAA 3893

Qy       3934 AGTTTCTGTCTCACTTT 3950
             |||||||||||||||||
Db       3894 AGTTTCTGTCTCACTTT 3910
```

```
RESULT 3
AF104260
LOCUS       AF104260               3362 bp    mRNA    linear   PRI 09-JAN-2002
DEFINITION  Homo sapiens HIWI mRNA, complete cds.
ACCESSION   AF104260
VERSION     AF104260.2  GI:18098557
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3362)
  AUTHORS   Cox,D.N., Chao,A., Baker,J., Chang,L., Qiao,D. and Lin,H.
  TITLE     A novel class of evolutionarily conserved genes defined by piwi are
            essential for stem cell self-renewal
  JOURNAL   Genes Dev. 12 (23), 3715-3727 (1998)
  MEDLINE   99069219
   PUBMED   9851978
REFERENCE   2  (bases 1 to 3362)
  AUTHORS   Qiao,D., Zeeman,A.-M., Deng,W., Looijenga,L.H.J. and Lin,H.
  TITLE     Molecular characterization of hiwi, a human member of the piwi stem
            cell gene family whose overexpression is correlated to seminomas
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 3362)
  AUTHORS   Lin,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1998) Cell Biology, Duke University Medical
            Center, 412 Nanaline Duke Bldg., Research Dr., Durham, NC 27710,
            USA
REFERENCE   4  (bases 1 to 3362)
  AUTHORS   Qiao,D., Zeeman,A.-M., Deng,W., Looijenga,L.H.J. and Lin,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Cell Biology, Duke University, 412 Nanaline
            Duke Building, DUMC, Durham, NC 27710, USA
  REMARK    Nucleotide sequence updated by submitter
COMMENT     On Jan 9, 2002 this sequence version replaced gi:4038412.
FEATURES             Location/Qualifiers
     source          1. .3362
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /map="12q24.33"
                     /tissue_type="testis"
     CDS             57. .2642
                     /note="member of the PIWI family"
                     /codon_start=1
                     /product="HIWI"
                     /protein_id="AAC97371.2"
                     /db_xref="GI:18098558"
                     /translation="MTGRARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEG
                     ELFGRGRQRGTAGGTAKSQGLQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVKE
                     SKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKC
                     HAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRR
                     LLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRS
                     ETVLDFMFNFYHQTEEHKFQEQVSKELIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKK
                     ADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGL
                     TDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLL
                     SFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA
                     NSLIQNLFKVTPAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKD
                     KYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLK
                     LVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAA
                     LRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVV
                     KKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIY
                     DNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSN
                     RLYYL"
BASE COUNT     1026 a    659 c    759 g    918 t
ORIGIN
```

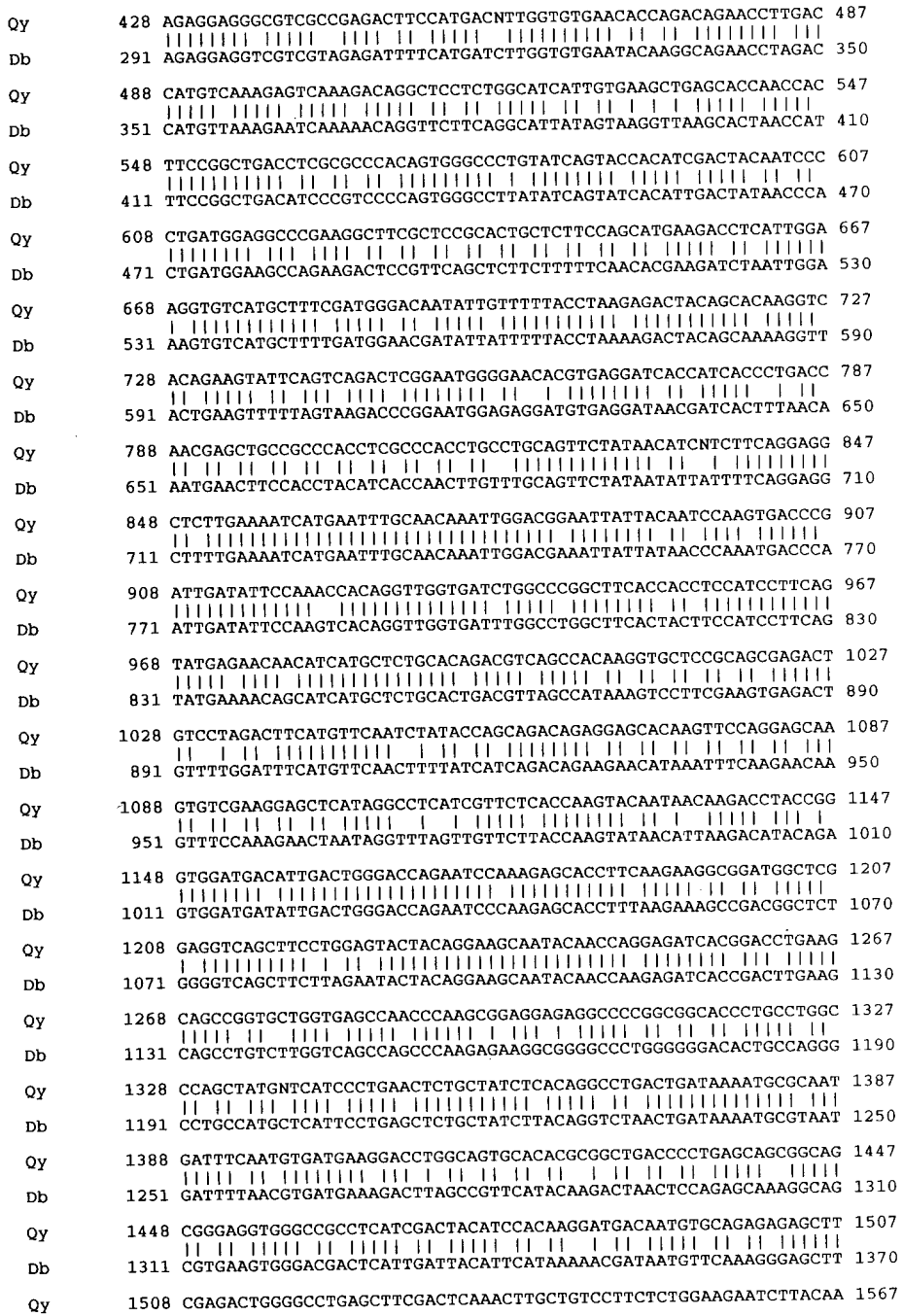
```
  Query Match           45.2%;  Score 1835;  DB 9;  Length 3362;
  Best Local Similarity  81.8%;  Pred. No. 0;
  Matches 2127;  Conservative    0;  Mismatches  469;  Indels    3;  Gaps    1;

Qy        188 AAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGAGACGGTG 247
              | |||||||||  ||||||||||  | ||| | || ||||| || ||||||||||| | |
Db         54 ACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCG 113

Qy        248 CAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACCTCAACAG 307
              ||||  || ||  |  |||| || |||||||||||| || || |  || || ||||| | |
Db        114 CAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCG 173

Qy        308 TCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGTCGGAGCC 367
              | ||  |||||||||| ||  ||||||| ||||| ||||| || |    ||  ||||
Db        174 CCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAA---CAGCAGGAGGA 230

Qy        368 ACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACTGGCAGAG 427
              ||| |||||||||||| ||| ||||| || ||||| ||||||||| | ||  | ||||||
Db        231 ACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTAGCAGAG 290
```


Qy 428 AGAGGAGGGCGTCGCCGAGACTTCCATGACNTTGGTGTGAACACCAGACAGAACCTTGAC 487
Db 291 AGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGAC 350
Qy 488 CATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCAC 547
Db 351 CATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCAT 410
Qy 548 TTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCC 607
Db 411 TTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCA 470
Qy 608 CTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGA 667
Db 471 CTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGA 530
Qy 668 AGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGTC 727
Db 531 AAGTGTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTT 590
Qy 728 ACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCACCCTGACC 787
Db 591 ACTGAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACA 650
Qy 788 AACGAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTTCAGGAGG 847
Db 651 AATGAACTTCCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGG 710
Qy 848 CTCTTGAAAATCATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAGTGACCCG 907
Db 711 CTTTTGAAAATCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCA 770
Qy 908 ATTGATATTCCAAACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCATCCTTCAG 967
Db 771 ATTGATATTCCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAG 830
Qy 968 TATGAGAACAACATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAGCGAGACT 1027
Db 831 TATGAAAACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACT 890
Qy 1028 GTCCTAGACTTCATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCAGGAGCAA 1087
Db 891 GTTTTGGATTTCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAA 950
Qy 1088 GTGTCGAAGGAGCTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGACCTACCGG 1147
Db 951 GTTTCCAAAGAACTAATAGGTTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGA 1010
Qy 1148 GTGGATGACATTGACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGATGGCTCG 1207
Db 1011 GTGGATGATATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCT 1070
Qy 1208 GAGGTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGACCTGAAG 1267
Db 1071 GGGGTCAGCTTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAG 1130
Qy 1268 CAGCCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCTGCCTGGC 1327
Db 1131 CAGCCTGTCTTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGG 1190
Qy 1328 CCAGCTATGNTCATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAATGCGCAAT 1387
Db 1191 CCTGCCATGCTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAAT 1250
Qy 1388 GATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCAGCGGCAG 1447
Db 1251 GATTTTAACGTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAG 1310
Qy 1448 CGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAGAGAGCTT 1507
Db 1311 CGTGAAGTGGGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTT 1370
Qy 1508 CGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAATCTTACAA 1567
Db 1371 CGAGACTGGGGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAA 1430
Qy 1568 TCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGCAGACTGG 1627
Db 1431 ACAGAAAAGATTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGG 1490
Qy 1628 TCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGGCTGCTG 1687
Db 1491 TCCAAAGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTG 1550
Qy 1688 ATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTTCAAAGTG 1747
Db 1551 ATCTATACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTT 1610
Qy 1748 ACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGACAGAACA 1807
Db 1611 ACACCAGCCATGGGCATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACT 1670
Qy 1808 GAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGT 1867
Db 1671 GAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGT 1730
Qy 1868 CTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTGTACAGAC 1927
Db 1731 CTGTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGAT 1790
Qy 1928 TGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATG 1987
Db 1791 TGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATG 1850
Qy 1988 GCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGG 2047
Db 1851 GCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGG 1910
Qy 2048 GTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACACCACA 2107
Db 1911 GTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACA 1970
Qy 2108 GCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGG 2167
Db 1971 GCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGG 2030
Qy 2168 TTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGC 2227
Db 2031 TTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGC 2090
Qy 2228 TTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATC 2287
Db 2091 CTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATC 2150
Qy 2288 GTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCA 2347
Db 2151 GTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCA 2210
Qy 2348 CAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATC 2407
Db 2211 CAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGACTAACGGTAATT 2270
Qy 2408 GTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAAC 2467
Db 2271 GTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAAT 2330
Qy 2468 CCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTC 2527
Db 2331 CCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTT 2390
Qy 2528 ATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTAT 2587
Db 2391 ATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTAT 2450
Qy 2588 GACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTG 2647

```
Db       2451 GACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATC 2510

Qy       2648 TACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTG 2707
             || || || ||||| || ||||| || || ||||| |||||||||||| || |||||||||
Db       2511 TATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTG 2570

Qy       2708 GCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTC 2767
             || || || || |||||||| || |||||||||||||| || || ||||| ||||||||
Db       2571 GCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTT 2630

Qy       2768 TACTACCTCTAACCAGCAG 2786
              ||||||||||||||| ||||
Db       2631 TACTACCTCTAACCTGCAG 2649


RESULT 4
BC028581
LOCUS       BC028581               3399 bp    mRNA    linear   PRI 25-APR-2002
DEFINITION  Homo sapiens, piwi-like 1 (Drosophila), clone MGC:26748
            IMAGE:4827445, mRNA, complete cds.
ACCESSION   BC028581
VERSION     BC028581.1  GI:20306781
KEYWORDS    MGC.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3399)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 34 Row: b Column: 20
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, Similarity but not identity to protein.
FEATURES             Location/Qualifiers
     source          1. .3399
                     /organism="Homo sapiens"
                     /db_xref="LocusID:9271"
                     /db_xref="taxon:9606"
                     /clone="MGC:26748 IMAGE:4827445"
                     /tissue_type="Testis"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
     CDS             74. .2659
                     /codon_start=1
                     /product="piwi-like 1 (Drosophila)"
                     /protein_id="AAH28581.1"
                     /db_xref="GI:20306782"
                     /translation="MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEG
                     ELFGRGRQRGTAGGTAKSQGLQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVKE
                     SKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKC
                     HAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRR
                     LLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRS
                     ETVLDFMFNFYHQTEEHKFQEQVSKELIGLVVLTKYNNKTYRVDDIDWDQNPKSTFKK
                     ADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGL
                     TDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLL
                     SFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVNPLDNWLLIYTRRNYEAA
                     NSLIQNLFKVTPAMGMQMKKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKD
                     KYDAIKKYPCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLK
                     LVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAA
                     LRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVV
                     KKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIY
                     DNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSN
                     RLYYL"
BASE COUNT     1048 a    662 c    766 g    923 t
ORIGIN

  Query Match           45.2%;  Score 1835;  DB 9;  Length 3399;
  Best Local Similarity 81.8%;  Pred. No. 0;
  Matches 2127;  Conservative    0;  Mismatches  469;  Indels    3;  Gaps    1;

Qy        188 AAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGAGACGGTG 247
              | |||||||||  ||||||||||  | ||| | || ||||| || |||||||||||| | |
Db         71 ACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCG 130

Qy        248 CAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACCTCAACAG 307
              ||||  || ||  |  |||| || |||||||||||| || || |  || || ||||| | |
Db        131 CAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCG 190

Qy        308 TCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGTCGGAGCC 367
              | ||  |||||||||| ||  ||||||| ||||| ||||| || |    ||  ||||
Db        191 CCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAA---CAGCAGGAGGA 247

Qy        368 ACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACTGGCAGAG 427
              ||| |||||||||||| ||| ||||| || ||||| |||||||||| | ||  | ||||||
Db        248 ACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTAGCAGAG 307

Qy        428 AGAGGAGGGCGTCGCCGAGACTTCCATGACNTTGGTGTGAACACCAGACAGAACCTTGAC 487
              |||||||| |||||  |||| || |||||  |||||||||| || || |||||||| |||
Db        308 AGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGAC 367

Qy        488 CATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCAC 547
              ||||| ||||| ||||| ||||| || || ||||| || || | | | ||||| |||||
Db        368 CATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCAT 427

Qy        548 TTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCC 607
              |||||||||| || || || |||||||||  | |||||||| ||||| |||| || ||
Db        428 TTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCA 487

Qy        608 CTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGA 667
              |||||||| ||| |||| || || || || || || || || || ||||| || ||||||
Db        488 CTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGA 547

Qy        668 AGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGTC 727
              | |||||||||||| ||||| || ||||| |||||||||| |||||||||| |||||
Db        548 AAGTGTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTT 607

Qy        728 ACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCACCCTGACC 787
              || ||||| || ||| |||| |||||||| ||  | |||||||| || |||||  | ||
Db        608 ACTGAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACA 667

Qy        788 AACGAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTTCAGGAGG 847
              || || || || || || || || || || ||  |||||||||||| ||  | |||||||||
Db        668 AATGAACTTCCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGG 727

Qy        848 CTCTTGAAAATCATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAGTGACCCG 907
              || ||||||||||||||||||||||||||||||||||| |||||||| || |||| ||||||
Db        728 CTTTTGAAAATCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCA 787

Qy        908 ATTGATATTCCAAACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCATCCTTCAG 967
              ||||||||||||  |||||||||||||| ||||| |||||||| || ||||||||||||
Db        788 ATTGATATTCCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAG 847

Qy        968 TATGAGAACAACATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAGCGAGACT 1027
              ||||| |||| |||||||||||||| ||||| ||||| || || || || || ||||||
Db        848 TATGAAAACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACT 907
```

```
Qy       1028 GTCCTAGACTTCATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCAGGAGCAA 1087
Db        908 GTTTTGGATTTCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAA 967

Qy       1088 GTGTCGAAGGAGCTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGACCTACCGG 1147
Db         968 GTTTCCAAAGAACTAATAGGTTTAGTTGTTCTTACCAAGTATAACAATAAGACATACAGA 1027

Qy       1148 GTGGATGACATTGACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGATGGCTCG 1207
Db       1028 GTGGATGATATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCT 1087

Qy       1208 GAGGTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGACCTGAAG 1267
Db       1088 GAAGTCAGCTTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAG 1147

Qy       1268 CAGCCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCTGCCTGGC 1327
Db       1148 CAGCCTGTCTTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGG 1207

Qy       1328 CCAGCTATGNTCATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAATGCGCAAT 1387
Db       1208 CCTGCCATGCTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAAT 1267

Qy       1388 GATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCAGCGGCAG 1447
Db       1268 GATTTTAACGTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAG 1327

Qy       1448 CGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAGAGAGCTT 1507
Db       1328 CGTGAAGTGGGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTT 1387

Qy       1508 CGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAATCTTACAA 1567
Db       1388 CGAGACTGGGGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAA 1447

Qy       1568 TCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGCAGACTGG 1627
Db       1448 ACAGAAAAGATTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGG 1507

Qy       1628 TCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGGCTGCTG 1687
Db       1508 TCCAAAGAAACAAGAGGTGCACCATTAATTAGTGTTAATCCACTAGATAACTGGCTGTTG 1567

Qy       1688 ATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTTCAAAGTG 1747
Db       1568 ATCTATACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTT 1627

Qy       1748 ACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGACAGAACA 1807
Db       1628 ACACCAGCCATGGGCATGCAAATGAAAAAAGCAATAATGATTGAAGTGGATGACAGAACT 1687

Qy       1808 GAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGT 1867
Db       1688 GAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGT 1747

Qy       1868 CTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTGTACAGAC 1927
Db       1748 CTGTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCCGTGTACAGAT 1807

Qy       1928 TGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATG 1987
Db       1808 TGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATG 1867

Qy       1988 GCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGG 2047
Db       1868 GCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGG 1927

Qy       2048 GTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACACCACA 2107
Db       1928 GTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACA 1987

Qy       2108 GCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGG 2167
Db       1988 GCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGG 2047

Qy       2168 TTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGC 2227
Db       2048 TTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGC 2107

Qy       2228 TTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATC 2287
Db       2108 CTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATC 2167

Qy       2288 GTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCA 2347
Db       2168 GTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCA 2227

Qy       2348 CAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATC 2407
Db       2228 CAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGACTAACGGTAATT 2287

Qy       2408 GTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAAC 2467
Db       2288 GTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAAT 2347

Qy       2468 CCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTC 2527
Db       2348 CCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTT 2407

Qy       2528 ATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTAT 2587
Db       2408 ATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTAT 2467

Qy       2588 GACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTG 2647
Db       2468 GACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATC 2527

Qy       2648 TACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTG 2707
Db       2528 TATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTG 2587

Qy       2708 GCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTC 2767
Db       2588 GCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTT 2647

Qy       2768 TACTACCTCTAACCAGCAG 2786
Db       2648 TACTACCTCTAACCTGCAG 2666
```

```
RESULT 5
AF387507
LOCUS       AF387507               3421 bp    mRNA    linear   PRI 01-JUL-2001
DEFINITION  Homo sapiens PIWI protein mRNA, complete cds.
ACCESSION   AF387507
VERSION     AF387507.1  GI:14579644
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3421)
  AUTHORS   Sha,J.H.
  TITLE     Cloning and identification of human piwi protein related to testis
            development
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3421)
  AUTHORS   Sha,J.H., Li,J.M. and Zhou,Z.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-2001) Key Lab of Reproductive Medicine, Nanjing
            Medical University, 140 Han Zhong Road, Nanjing Medical University,
            Jiangsu 210029, China
FEATURES             Location/Qualifiers
     source          1. .3421
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="testis"
                     /dev_stage="adult"
     CDS             64. .2649
                     /note="higher expression in adult testis than 6 month old
                     embryo testis"
                     /codon_start=1
                     /product="PIWI protein"
                     /protein_id="AAK69348.1"
                     /db_xref="GI:14579645"
                     /translation="MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEG
                     ELFGRGRQRGTAGGTAKSQGLQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVKE
                     SKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKC
                     HAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRR
                     LLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRS
                     ETVLDFMFNFYHQTEEHKFQEQVSKELIGLVVLTKYNNKTYRVDDIDWDQNPKSTFKK
                     ADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGL
                     TDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLL
                     SFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA
                     NSLIQNLFKVTPAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKD
                     KYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLK
                     LVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAA
                     LRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVV
                     KKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIY
                     DNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSN
                     RLYYL"
     polyA_signal    3354. .3359
BASE COUNT     1071 a    661 c    767 g    922 t
ORIGIN
```

Query Match 45.2%; Score 1835; DB 9; Length 3421;
Best Local Similarity 81.8%; Pred. No. 0;
Matches 2127; Conservative 0; Mismatches 469; Indels 3; Gaps 1;

```
Qy        188 AAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGAGACGGTG 247
Db         61 ACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCG 120

Qy        248 CAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACCTCAACAG 307
Db        121 CAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCG 180

Qy        308 TCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGTCGGAGCC 367
Db        181 CCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAA---CAGCAGGAGGA 237

Qy        368 ACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACTGGCAGAG 427
Db        238 ACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTAGCAGAG 297

Qy        428 AGAGGAGGGCGTCGCCGAGACTTCCATGACNTTGGTGTGAACACCAGACAGAACCTTGAC 487
Db        298 AGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGAC 357

Qy        488 CATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCAC 547
Db        358 CATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCAT 417

Qy        548 TTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCC 607
Db        418 TTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCA 477

Qy        608 CTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGA 667
Db        478 CTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGA 537

Qy        668 AGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGTC 727
Db        538 AAGTGCCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTT 597

Qy        728 ACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCACCCTGACC 787
Db        598 ACTGAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACA 657

Qy        788 AACGAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTTCAGGAGG 847
Db        658 AATGAACTTCCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGG 717

Qy        848 CTCTTGAAAATCATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAGTGACCCG 907
Db        718 CTTTTGAAAATCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCA 777

Qy        908 ATTGATATTCCAAACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCATCCTTCAG 967
Db        778 ATTGATATTCCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAG 837

Qy        968 TATGAGAACAACATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAGCGAGACT 1027
Db        838 TATGAAAACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACT 897

Qy       1028 GTCCTAGACTTCATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCAGGAGCAA 1087
Db        898 GTTTTGGATTTCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAA 957

Qy       1088 GTGTCGAAGGAGCTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGACCTACCGG 1147
Db        958 GTTTCCAAAGAACTAATAGGTTTAGTTGTTCTTACCAAGTATAACAATAAGACATACAGA 1017

Qy       1148 GTGGATGACATTGACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGATGGCTCG 1207
Db       1018 GTGGATGATATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCT 1077

Qy       1208 GAGGTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGACCTGAAG 1267
Db       1078 GAAGTCAGCTTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAG 1137

Qy       1268 CAGCCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCTGCCTGGC 1327
Db       1138 CAGCCTGTCTTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGG 1197

Qy       1328 CCAGCTATGNTCATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAATGCGCAAT 1387
Db       1198 CCTGCCATGCTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAAT 1257

Qy       1388 GATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCAGCGGCAG 1447
Db       1258 GATTTTAACGTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAG 1317

Qy       1448 CGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAGAGAGCTT 1507
Db       1318 CGTGAAGTGGGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTT 1377

Qy       1508 CGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAATCTTACAA 1567
Db       1378 CGAGACTGGGGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAA 1437

Qy       1568 TCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGCAGACTGG 1627
Db       1438 ACAGAAAAGATTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGG 1497

Qy       1628 TCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGGCTGCTG 1687
Db       1498 TCCAAAGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTG 1557

Qy       1688 ATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTTCAAAGTG 1747
Db       1558 ATCTATACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTT 1617

Qy       1748 ACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGACAGAACA 1807
Db       1618 ACACCAGCCATGGGCATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACT 1677

Qy       1808 GAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGT 1867
Db       1678 GAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGT 1737
```

```
Qy       1868 CTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTGTACAGAC 1927
             || |||||||||||||||||||||||||| ||||| || || || ||| ||||||||||
Db       1738 CTGTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGAT 1797

Qy       1928 TGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATG 1987
             ||||| |||||||||||||||||||||||| ||| | ||||| |||||||| ||||||
Db       1798 TGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATG 1857

Qy       1988 GCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGG 2047
             |||||||| || ||||| ||||| |||||||||||||||||||||||| ||||||||| ||
Db       1858 GCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGG 1917

Qy       2048 GTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACACCACA 2107
             |||||||| ||||||| || |  |||||||| |||||||| |||||||||||||  |||
Db       1918 GTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACA 1977

Qy       2108 GCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGG 2167
             |||||||||||||| |||||||||| || ||||||||||||||||||||||||||||||
Db       1978 GCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGG 2037

Qy       2168 TTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGC 2227
             ||||| |||||| | |||||||| | || |||||||||||| ||||| ||||| || |||
Db       2038 TTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGC 2097

Qy       2228 TTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATC 2287
              ||||||| |||||||||||||||| | ||||||||| ||||||||||||||  |||||
Db       2098 CTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATC 2157

Qy       2288 GTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCA 2347
             |||||||| || || || || ||||| |||||||| || |||||  || || || || |||
Db       2158 GTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCA 2217

Qy       2348 CAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATC 2407
             |||||  | ||||| || || ||  | || |||||||||||||| ||||| || |||||
Db       2218 CAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGACTAACGGTAATT 2277

Qy       2408 GTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAAC 2467
             ||||||||||  | || ||  |||| |||||||||||||||| ||||||||||||||
Db       2278 GTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAAT 2337

Qy       2468 CCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTC 2527
             || ||||| || ||||| || ||||| || || |||||||||| ||||||||||||||
Db       2338 CCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTT 2397

Qy       2528 ATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTAT 2587
             |||||||||||||| |||||||| || ||||| || || ||||| ||||||||||||||||
Db       2398 ATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTAT 2457

Qy       2588 GACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTG 2647
             |||| ||| |||||||||||| |||||||| ||||| |||| |||||| | |||||| |
Db       2458 GACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATC 2517

Qy       2648 TACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTG 2707
             || || || ||||| || ||||| || || ||||| |||||||||| || |||||||||
Db       2518 TATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTG 2577

Qy       2708 GCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTC 2767
             || || || || |||||||| || |||||||||||||| || || ||||| ||||||||
Db       2578 GCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTT 2637

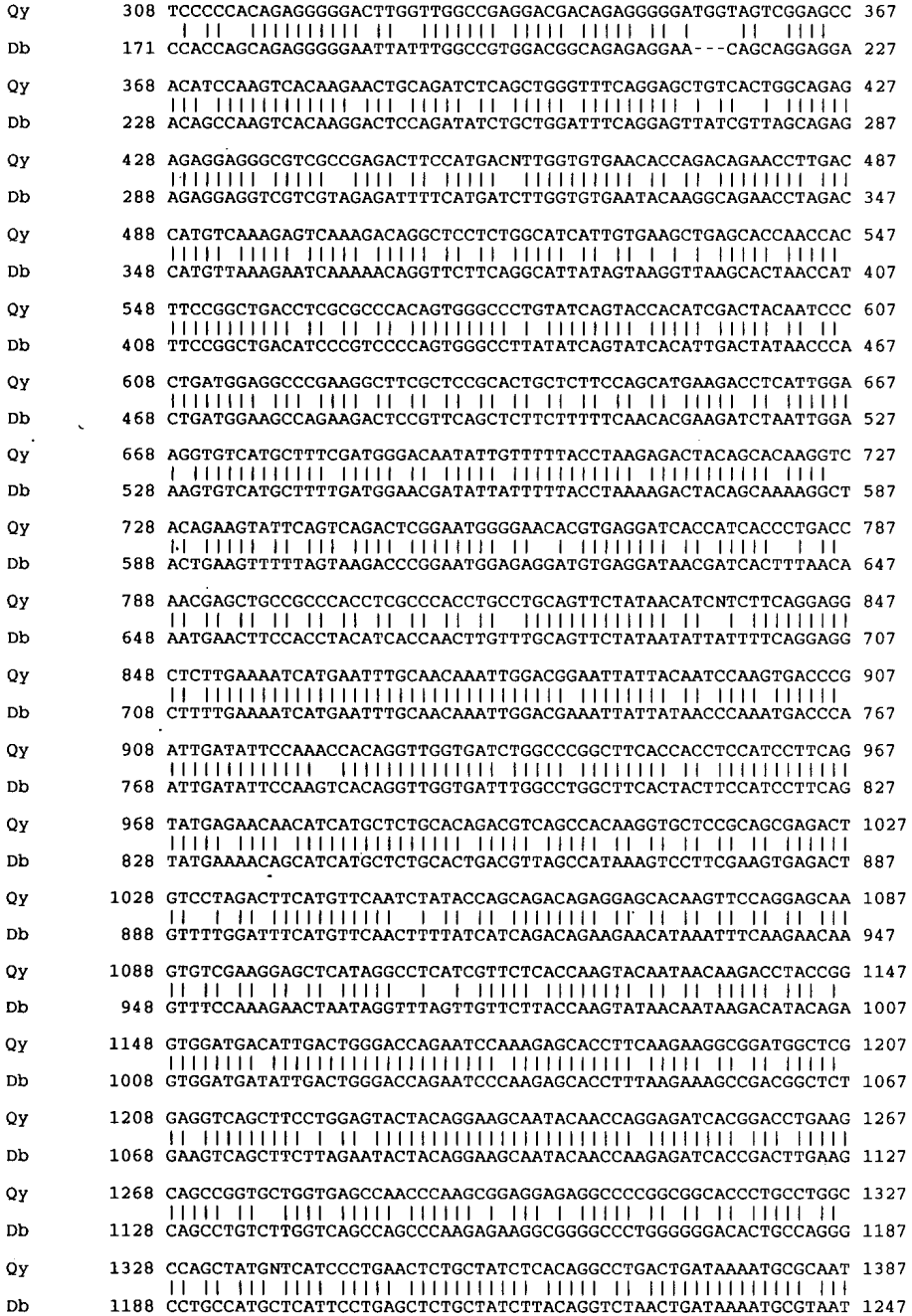
Qy       2768 TACTACCTCTAACCAGCAG 2786
              |||||||||||||| ||||
Db       2638 TACTACCTCTAACCTGCAG 2656
```

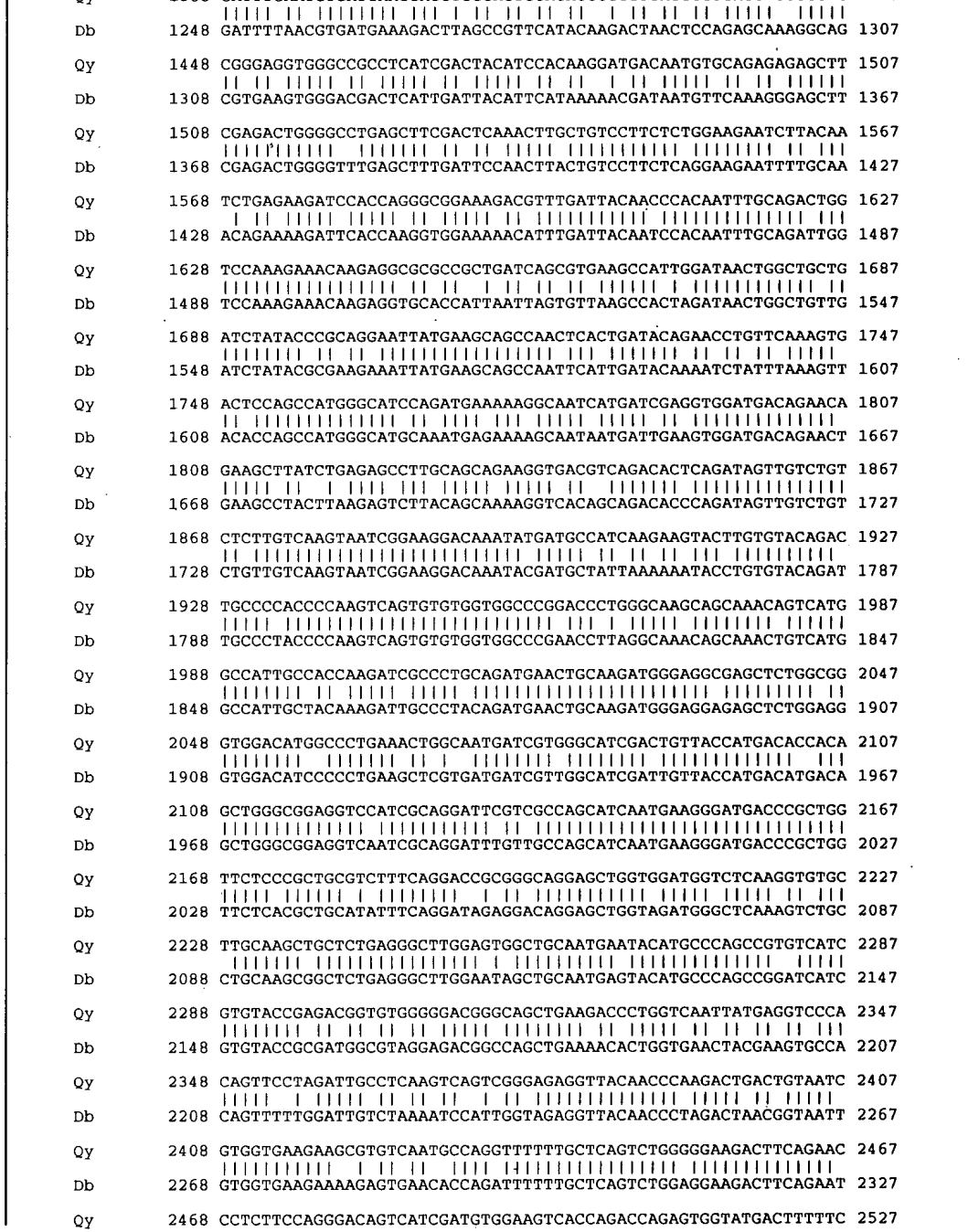
```
RESULT 6
AK093133
LOCUS       AK093133               2579 bp    mRNA    linear   PRI 15-JUL-2002
DEFINITION  Homo sapiens cDNA FLJ35814 fis, clone TESTI2006083, highly similar
            to Mus musculus mRNA for MIWI (piwi).
ACCESSION   AK093133
VERSION     AK093133.1  GI:21751907
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens testis cDNA to mRNA, clone_lib:TESTI2
            clone:TESTI2006083.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,
            Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R.,
            Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kawakami,B.,
            Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and
            Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2579)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1. .2579
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="TESTI2006083"
                     /tissue_type="testis"
                     /clone_lib="TESTI2"
                     /note="cloning vector: pME18SFL3"
     CDS             54. .2543
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAC04068.1"
                     /db_xref="GI:21751908"
                     /translation="MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEG
                     ELFGRGRQRGTAGGTAKSQGLQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVKE
                     SKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKC
                     HAFDGTILFLPKRLQQKATEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRR
                     LLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRS
                     ETVLDFMFNFYHQTEEHKFQEQVSKELIGLVVLTKYNNKTYRVDDIDWDQNPKSTFKK
                     ADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGL
                     TDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLL
                     SFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA
                     NSLIQNLFKVTPAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKD
                     KYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLK
                     LVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAA
                     LRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVV
                     KKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIY
                     DNSGLKPDHIQRLTYKLCHIYYNWPVSASTC"
BASE COUNT      797 a    542 c    613 g    627 t
ORIGIN

  Query Match             42.9%;  Score 1742.8;  DB 9;  Length 2579;
  Best Local Similarity   81.7%;  Pred. No. 6.3e-308;
  Matches 2024;  Conservative    0;  Mismatches  451;  Indels    3;  Gaps    1;

Qy        188 AAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGAGACGGTG 247
              | |||||||||  ||||||||||  | ||| | || ||||| || ||||||||||| | |
Db         51 ACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCG 110

Qy        248 CAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACCTCAACAG 307
              ||||  || ||  |  |||| || |||||||||||| || || |  || || ||||| | |
Db        111 CAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCG 170
```





```
                || ||||| || ||||| || ||||| || || ||||||||||| ||||||||||||||
Db       2328 CCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTT 2387

Qy       2528 ATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTAT 2587
                ||||||||||||||| |||||||| || ||||| || || ||||| ||||||||||||||
Db       2388 ATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTAT 2447

Qy       2588 GACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTG 2647
                |||| ||| |||||||||||| |||||||| |||||  |||| |||||| | ||||||
Db       2448 GACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATC 2507

Qy       2648 TACTATAATTGGCCTGGA 2665
                || || || ||||| | |
Db       2508 TATTACAACTGGCCAGTA 2525


RESULT 7
AF264004
LOCUS       AF264004                2328 bp    mRNA    linear   PRI 21-AUG-2001
DEFINITION  Homo sapiens HIWI mRNA, complete cds.
ACCESSION   AF264004
VERSION     AF264004.1  GI:15216446
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2328)
  AUTHORS   Sharma,A.K., Nelson,M.C., Brandt,J.E., Wessman,M., Muhmud,N.,
            Weller,K.P. and Hoffman,R.
  TITLE     Human CD34+ Stem Cells Express the hiwi Gene, a Human Homolog of
            the Drosophila Gene piwi
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2328)
  AUTHORS   Sharma,A.K. and Hoffman,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-2000) Medicine, UIC, 900 South Ashland Avenue,
            Chicago, IL 60607, USA
FEATURES             Location/Qualifiers
     source          1. .2328
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="testis"
     CDS             1. .2328
                     /codon_start=1
                     /product="HIWI"
                     /protein_id="AAK92281.1"
                     /db_xref="GI:15216447"
                     /translation="MIFGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALY
                     QYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTEVFSKTRNG
                     EDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRL
                     VIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKELI
                     GLVVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVL
                     VSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQRE
                     VGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADW
                     SKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDD
                     RTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQ
                     QTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASIN
                     EGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLK
                     TLVNYEVPQFLDCLKSIGRGYNPRLTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVE
                     VTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVI
                     RVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL"
BASE COUNT      716 a    496 c    520 g    596 t
ORIGIN

  Query Match           41.4%;  Score 1680.6;  DB 9;  Length 2328;
  Best Local Similarity  82.7%;  Pred. No. 1.4e-296;
  Matches 1920;  Conservative    0;  Mismatches  403;  Indels    0;  Gaps    0;

Qy        457 CNTTGGTGTGAACACCAGACAGAACCTTGACCATGTCAAAGAGTCAAAGACAGGCTCCTC 516
              | |||||||||||| || |||||| | |||||||| ||||| ||||| ||||| || ||
Db          6 CTTTGGTGTGAACACAAGGCAGAACTTAGACCATGTTAAAGAATCAAAAACAGGTTCTTC 65

Qy        517 TGGCATCATTGTGAAGCTGAGCACCAACCACTTCCGGCTGACCTCGCGCCCACAGTGGGC 576
              ||||| || || | | | ||||| ||||| |||||||||||| || || || ||||||||
Db         66 AGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGACATCCCGTCCCCAGTGGGC 125

Qy        577 CCTGTATCAGTACCACATCGACTACAATCCCCTGATGGAGGCCCGAAGGCTTCGCTCCGC 636
              | | |||||||| ||||| ||||| || || |||||||| ||| |||| || || || ||
Db        126 CTTATATCAGTATCACATTGACTATAACCCACTGATGGAAGCCAGAAGACTCCGTTCAGC 185

Qy        637 ACTGCTCTTCCAGCATGAAGACCTCATTGGAAGGTGTCATGCTTTCGATGGGACAATATT 696
              || || || || || ||||| || ||||||| |||||||||||| ||||| || |||||
Db        186 TCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCATGCTTTTGATGGAACGATATT 245

Qy        697 GTTTTTACCTAAGAGACTACAGCACAAGGTCACAGAAGTATTCAGTCAGACTCGGAATGG 756
              |||||||||||| |||||||||| ||||| || ||||| || ||| |||| ||||||||
Db        246 ATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTTTTTAGTAAGACCCGGAATGG 305

Qy        757 GGAACACGTGAGGATCACCATCACCCTGACCAACGAGCTGCCGCCCACCTCGCCCACCTG 816
              ||  | |||||||| || |||||  | || || || || || || || || || || ||
Db        306 AGAGGATGTGAGGATAACGATCACTTTAACAAATGAACTTCCACCTACATCACCAACTTG 365

Qy        817 CCTGCAGTTCTATAACATCNTCTTCAGGAGGCTCTTGAAAATCATGAATTTGCAACAAAT 876
              |||||||||||| ||  | |||||||||| |||||||||||||||||||||||||||||
Db        366 TTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAAATCATGAATTTGCAACAAAT 425

Qy        877 TGGACGGAATTATTACAATCCAAGTGACCCGATTGATATTCCAAACCACAGGTTGGTGAT 936
              |||||| |||||||| || |||| |||||| |||||||||||||  ||||||||||||||
Db        426 TGGACGAAATTATTATAACCCAAATGACCCAATTGATATTCCAAGTCACAGGTTGGTGAT 485

Qy        937 CTGGCCCGGCTTCACCACCTCCATCCTTCAGTATGAGAACAACATCATGCTCTGCACAGA 996
              ||||| |||||||| || ||||||||||||||||| |||| ||||||||||||||| ||
Db        486 TTGGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAACAGCATCATGCTCTGCACTGA 545

Qy        997 CGTCAGCCACAAGGTGCTCCGCAGCGAGACTGTCCTAGACTTCATGTTCAATCTATACCA 1056
              ||| ||||| || || || || || |||||||  | || ||||||||||  | || ||
Db        546 CGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGATTTCATGTTCAACTTTTATCA 605

Qy       1057 GCAGACAGAGGAGCACAAGTTCCAGGAGCAAGTGTCGAAGGAGCTCATAGGCCTCATCGT 1116
              |||||||| || || || || || || ||||| || || || || |||||  |  | ||
Db        606 TCAGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAAGAACTAATAGGTTTAGTTGT 665

Qy       1117 TCTCACCAAGTACAATAACAAGACCTACCGGGTGGATGACATTGACTGGGACCAGAATCC 1176
              ||| |||||||| || || ||||| ||| | |||||||| ||||||||||||||||||||
Db        666 TCTTACCAAGTATAACAATAAGACATACAGAGTGGATGATATTGACTGGGACCAGAATCC 725

Qy       1177 AAAGAGCACCTTCAAGAAGGCGGATGGCTCGGAGGTCAGCTTCCTGGAGTACTACAGGAA 1236
              |||||||||| ||||| || || ||||| || ||||||||| | || ||||||||||
Db        726 CAAGAGCACCTTTAAGAAAGCCGACGGCTCTGAAGTCAGCTTCTTAGAATACTACAGGAA 785

Qy       1237 GCAATACAACCAGGAGATCACGGACCTGAAGCAGCCGGTGCTGGTGAGCCAACCCAAGCG 1296
              |||||||||||| |||||||| ||| ||||||||| ||  |||| ||||| |||||| |
Db        786 GCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGCCAGCCCAAGAG 845

Qy       1297 GAGGAGAGGCCCCGGCGGCACCCTGCCTGGCCCAGCTATGNTCATCCCTGAACTCTGCTA 1356
              ||| | ||||| || || || ||||| || || || ||| |||| ||||| ||||||||
Db        846 AAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATGCTCATTCCTGAGCTCTGCTA 905

Qy       1357 TCTCACAGGCCTGACTGATAAAATGCGCAATGATTTCAATGTGATGAAGGACCTGGCAGT 1416
              ||| ||||| || |||||||||||| |||||||| || |||||||| ||| | || ||
Db        906 TCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAACGTGATGAAAGACTTAGCCGT 965

Qy       1417 GCACACGCGGCTGACCCCTGAGCAGCGGCAGCGGGAGGTGGGCCGCCTCATCGACTACAT 1476
              || ||  | || || || ||||| ||||||| || ||||| || ||||| || ||||||
Db        966 TCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTGGGACGACTCATTGATTACAT 1025

Qy       1477 CCACAAGGATGACAATGTGCAGAGAGAGCTTCGAGACTGGGGCCTGAGCTTCGACTCAAA 1536
              || ||  | || ||||| || || |||||||||||||||||  ||||||| || || ||
Db       1026 TCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGGGTTTGAGCTTTGATTCCAA 1085

Qy       1537 CTTGCTGTCCTTCTCTGGAAGAATCTTACAATCTGAGAAGATCCACCAGGGCGGAAAGAC 1596
              ||| |||||||||| |||||||| || ||| | || ||||| ||||| || ||||| ||
Db       1086 CTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAGATTCACCAAGGTGGAAAAAC 1145

Qy       1597 GTTTGATTACAACCCACAATTTGCAGACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGAT 1656
```

```
Db       1146 ATTTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAACAAGAGGTGCACCATTAAT 1205

Qy       1657 CAGCGTGAAGCCATTGGATAACTGGCTGCTGATCTATACCCGCAGGAATTATGAAGCAGC 1716

Db       1206 TAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACGCGAAGAAATTATGAAGCAGC 1265

Qy       1717 CAACTCACTGATACAGAACCTGTTCAAAGTGACTCCAGCCATGGGCATCCAGATGAAAAA 1776

Db       1266 CAATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGCATGCAAATGAGAAA 1325

Qy       1777 GGCAATCATGATCGAGGTGGATGACAGAACAGAAGCTTATCTGAGAGCCTTGCAGCAGAA 1836

Db       1326 AGCAATAATGATTGAAGTGGATGACAGAACTGAAGCCTACTTAAGAGTCTTACAGCAAAA 1385

Qy       1837 GGTGACGTCAGACACTCAGATAGTTGTCTGTCTCTTGTCAAGTAATCGGAAGGACAAATA 1896

Db       1386 GGTCACAGCAGACACCCAGATAGTTGTCTGTCTGTTGTCAAGTAATCGGAAGGACAAATA 1445

Qy       1897 TGATGCCATCAAGAAGTACTTGTGTACAGACTGCCCCACCCCAAGTCAGTGTGTGGTGGC 1956

Db       1446 CGATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGC 1505

Qy       1957 CCGGACCCTGGGCAAGCAGCAAACAGTCATGGCCATTGCCACCAAGATCGCCCTGCAGAT 2016

Db       1506 CCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAAAGATTGCCCTACAGAT 1565

Qy       2017 GAACTGCAAGATGGGAGGCGAGCTCTGGCGGGTGGACATGGCCCTGAAACTGGCAATGAT 2076

Db       1566 GAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCTCGTGATGAT 1625

Qy       2077 CGTGGGCATCGACTGTTACCATGACACCACAGCTGGGCGGAGGTCCATCGCAGGATTCGT 2136

Db       1626 CGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGCAGGATTTGT 1685

Qy       2137 CGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCGG 2196

Db       1686 TGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCATATTTCAGGATAGAGG 1745

Qy       2197 GCAGGAGCTGGTGGATGGTCTCAAGGTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGG 2256

Db       1746 ACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGGCTCTGAGGGCTTGGAATAG 1805

Qy       2257 CTGCAATGAATACATGCCCAGCCGTGTCATCGTGTACCGAGACGGTGTGGGGGACGGGCA 2316

Db       1806 CTGCAATGAGTACATGCCCAGCCGGATCATCGTGTACCGCGATGGCGTAGGAGACGGCCA 1865

Qy       2317 GCTGAAGACCCTGGTCAATTATGAGGTCCCACAGTTCCTAGATTGCCTCAAGTCAGTCGG 2376

Db       1866 GCTGAAAACACTGGTGAACTACGAAGTGCCACAGTTTTTGGATTGTCTAAAATCCATTGG 1925

Qy       2377 GAGAGGTTACAACCCAAGACTGACTGTAATCGTGGTGAAGAAGCGTGTCAATGCCAGGTT 2436

Db       1926 TAGAGGTTACAACCCTAGACTAACGGTAATTGTGGTGAAGAAAAGAGTGAACACCAGATT 1985

Qy       2437 TTTTGCTCAGTCTGGGGGAAGACTTCAGAACCCTCTTCCAGGGACAGTCATCGATGTGGA 2496

Db       1986 TTTTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTCCTGGAACAGTTATTGATGTAGA 2045

Qy       2497 AGTCACCAGACCAGAGTGGTATGACTTTTTCATCGTGAGCCAGGCAGTGAGAAGCGGGAG 2556

Db       2046 GGTTACCAGACCAGAATGGTATGACTTTTTTATCGTGAGCCAGGCTGTGAGAAGTGGTAG 2105

Qy       2557 TGTGTCCCCAACACACTACAATGTCATCTATGACAGCAGTGGCCTGAAGCCCGACCACAT 2616

Db       2106 TGTTTCTCCCACACATTACAATGTCATCTATGACAACAGCGGCCTGAAGCCAGACCACAT 2165

Qy       2617 CCAGCGGCTGACATACAAGNTCTGCCACGTGTACTATAATTGGCCTGGAGTCATCCGAGT 2676

Db       2166 ACAGCGCTTGACCTACAAGCTGTGCCACATCTATTACAACTGGCCAGGTGTCATTCGTGT 2225

Qy       2677 CCCTGCACCTTGCCAGTATGCACACAAGCTGGCCTTCCTCGTGGGCCAGAGCATCCACAG 2736

Db       2226 TCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTGTTGGCCAGAGTATTCACAG 2285

Qy       2737 AGAGCCAAACCTCTCCCTGTCCAACCGCCTCTACTACCTCTAA 2779

Db       2286 AGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAA 2328
```

RESULT 8
AC111089

```
LOCUS       AC111089               236516 bp    DNA     linear   HTG 15-MAY-2002
DEFINITION  Mus musculus clone RP23-156F20, WORKING DRAFT SEQUENCE, 43 ordered
            pieces.
ACCESSION   AC111089
VERSION     AC111089.4  GI:20800211
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 236516)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-156F20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 236516)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 236516)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
```

```
JOURNAL   Submitted (15-MAY-2002) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On May 15, 2002 this sequence version replaced gi:20531907.
          All repeats were identified using RepeatMasker:
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          -------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
          -------------- Project Information
              Center project name: L22447
              Center clone name: 156_F_20
          -------------- Summary Statistics
              Sequencing vector: Plasmid; n/a; 100% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.960731
              Consensus quality: 223901 bases at least Q40
              Consensus quality: 229385 bases at least Q30
              Consensus quality: 231373 bases at least Q20
              Insert size: 215000; agarose-fp
              Insert size: 232316; sum-of-contigs
              Quality coverage: 9.5 in Q20 bases; agarose-fp
              Quality coverage: 8.8 in Q20 bases; sum-of-contigs
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 43 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *        1      832: contig of 832 bp in length
          *      833  932: gap of      100 bp
          *      933     1688: contig of 756 bp in length
          *     1689 1788: gap of      100 bp
          *     1789     2478: contig of 690 bp in length
          *     2479 2578: gap of      100 bp
          *     2579     3250: contig of 672 bp in length
          *     3251 3350: gap of      100 bp
          *     3351     3997: contig of 647 bp in length
          *     3998 4097: gap of      100 bp
          *     4098     4783: contig of 686 bp in length
          *     4784 4883: gap of      100 bp
          *     4884     5701: contig of 818 bp in length
          *     5702 5801: gap of      100 bp
          *     5802     6065: contig of 264 bp in length
          *     6066 6165: gap of      100 bp
          *     6166     6942: contig of 777 bp in length
          *     6943 7042: gap of      100 bp
          *     7043     7881: contig of 839 bp in length
          *     7882 7981: gap of      100 bp
          *     7982     8774: contig of 793 bp in length
          *     8775 8874: gap of      100 bp
          *     8875     9779: contig of 905 bp in length
          *     9780 9879: gap of      100 bp
          *     9880    10662: contig of 783 bp in length
          *    10663 10762: gap of      100 bp
          *    10763    12006: contig of 1244 bp in length
          *    12007 12106: gap of      100 bp
          *    12107    13181: contig of 1075 bp in length
          *    13182 13281: gap of      100 bp
          *    13282    13992: contig of 711 bp in length
          *    13993 14092: gap of      100 bp
          *    14093    14920: contig of 828 bp in length
          *    14921 15020: gap of      100 bp
          *    15021    15784: contig of 764 bp in length
          *    15785 15884: gap of      100 bp
          *    15885    16510: contig of 626 bp in length
          *    16511 16610: gap of      100 bp
          *    16611    17459: contig of 849 bp in length
          *    17460 17559: gap of      100 bp
          *    17560    18572: contig of 1013 bp in length
          *    18573 18672: gap of      100 bp
          *    18673    20236: contig of 1564 bp in length
          *    20237 20336: gap of      100 bp
          *    20337    21069: contig of 733 bp in length
          *    21070 21169: gap of      100 bp
          *    21170    22008: contig of 839 bp in length
          *    22009 22108: gap of      100 bp
          *    22109    23330: contig of 1222 bp in length
          *    23331 23430: gap of      100 bp
          *    23431    24828: contig of 1398 bp in length
          *    24829 24928: gap of      100 bp
          *    24929    26080: contig of 1152 bp in length
          *    26081 26180: gap of      100 bp
          *    26181    28066: contig of 1886 bp in length
          *    28067 28166: gap of      100 bp
          *    28167    29750: contig of 1584 bp in length
          *    29751 29850: gap of      100 bp
          *    29851    31180: contig of 1330 bp in length
          *    31181 31280: gap of      100 bp
          *    31281    34027: contig of 2747 bp in length
          *    34028 34127: gap of      100 bp
          *    34128    35895: contig of 1768 bp in length
          *    35896 35995: gap of      100 bp
          *    35996    38780: contig of 2785 bp in length
          *    38781 38880: gap of      100 bp
          *    38881    41448: contig of 2568 bp in length
          *    41449 41548: gap of      100 bp
          *    41549    43480: contig of 1932 bp in length
          *    43481 43580: gap of      100 bp
          *    43581    45553: contig of 1973 bp in length
          *    45554 45653: gap of      100 bp
          *    45654    49698: contig of 4045 bp in length
          *    49699 49798: gap of      100 bp
          *    49799   123340: contig of 73542 bp in length
          *   123341 123440: gap of      100 bp
          *   123441   140631: contig of 17191 bp in length
          *   140632 140731: gap of      100 bp
          *   140732   161772: contig of 21041 bp in length
          *   161773 161872: gap of      100 bp
          *   161873   193172: contig of 31300 bp in length
          *   193173 193272: gap of      100 bp
          *   193273   223340: contig of 30068 bp in length
          *   223341 223440: gap of      100 bp
          *   223441   236516: contig of 13076 bp in length.
FEATURES             Location/Qualifiers
     source          1. .236516
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-156F20"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .832
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    933. .1688
                     /note="assembly_fragment"
     misc_feature    1789. .2478
                     /note="assembly_fragment"
     misc_feature    2579. .3250
                     /note="assembly_fragment"

  Query Match             31.7%;  Score 1288.6;  DB 2;  Length 236516;
  Best Local Similarity   92.7%;  Pred. No. 5e-225;
  Matches 1429;  Conservative    1;  Mismatches    1;  Indels  111;  Gaps    3;

Qy       2514 GGTATGACTTTTTCATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACT 2573
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     218525 GGTATGACTTTTTCATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACT 218584

Qy       2574 ACAATGTCATCTATGACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACA 2633
```

```
Db    218585 ACAATGTCATCTATGACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACA 218644

Qy      2634 AGNTCTGCCACGTGTACTATAATTGGCCT------------------------------ 2662
             || ||||||||||||||||||||||||||
Db    218645 AGCTCTGCCACGTGTACTATAATTGGCCTGTGAGTGTCCTGCTATGGACCACCTACCCTG 218704

Qy      2663 ------------------------------------------------------------ 2662

Db    218705 GTTAGCATGCTGTGGTAGGGACCCTGGGAGTCCCGCCCAACTCAGCCCAGTGACTGGCCC 218764

Qy      2663 ------------GGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTGGC 2709
                         ||||||||||||||||||||||||||||||||||||||||||||||
Db    218765 TCTTTCCCCACAGGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTGGC 218824

Qy      2710 CTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTCTA 2769
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    218825 CTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTCTA 218884

Qy      2770 CTACCTCTAACCAGCAGCGAGGGCCGGCATGCTGGTTTCCTTCAGAGCTCTGCCCAATCC 2829
             |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    218885 CTACCTCTAACCAGCAGCGAGGGCCGGCATGCTGGTTTCCTTCAGAGCTCTG-CCAATCC 218943

Qy      2830 TTCAGTTCTATTTTTCATAGGATGAGGTTCTAGCGGGATGTTGGGGGTAGGGGTTGGGGG 2889
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    218944 TTCAGTTCTATTTTTCATAGGATGAGGTTCTAGCGGGATGTTGGGGGTAGGGGTTGGGGG 219003

Qy      2890 CCGCTGTTTTGTTTTGTTTTGTTTTTAGTTGTTGGTGTGTAGCGGCTATAGAGTTAACTA 2949
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219004 CCGCTGTTTTGTTTTGTTTTGTTTTTAGTTGTTGGTGTGTAGCGGCTATAGAGTTAACTA 219063

Qy      2950 AGGCTTATGTTTGCTTTGTTTCTCATGGTGTGTTGTGACAGTTTGCTTGGCAGAGTGGTG 3009
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219064 AGGCTTATGTTTGCTTTGTTTCTCATGGTGTGTTGTGACAGTTTGCTTGGCAGAGTGGTG 219123

Qy      3010 GCAAGGGCCGTGGTACCACAGATAGACTTTCCTAAAGCAGCCGTGTTGGCAGGTTCTTGG 3069
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219124 GCAAGGGCCGTGGTACCACAGATAGACTTTCCTAAAGCAGCCGTGTTGGCAGGTTCTTGG 219183

Qy      3070 AGGTGCTCCTGGGAGCACCTCCTATTGTACGCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 3129
             |||||||||||||||||||||||||||||||      ||||||||||||||||||||
Db    219184 AGGTGCTCCTGGGAGCACCTCCTATTGTACGCTG------CTCTCTCTCTCTCTCTCTCT 219237

Qy      3130 CTCTCTCTCTCTCTCTCTCTCTCTCTCTGCATGTTTGAATTTAAAAGAGGAAAGGACACA 3189
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219238 CTCTCTCTCTCTCTCTCTCTCTCTCTGCATGTTTGAATTTAAAAGAGGAAAGGACACA 219297

Qy      3190 AGGCAAGAATGCCCAGAGCATCCTAGATGGGTGCTATATGGGACTCTGACACCTCAGGTT 3249
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219298 AGGCAAGAATGCCCAGAGCATCCTAGATGGGTGCTATATGGGACTCTGACACCTCAGGTT 219357

Qy      3250 CTTAAAGATAAAAAGGCCAGAGGTACCTTTCTGTAATTCTTTTAAGAGTTAGGTCATAAG 3309
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219358 CTTAAAGATAAAAAGGCCAGAGGTACCTTTCTGTAATTCTTTTAAGAGTTAGGTCATAAG 219417

Qy      3310 GAATGAAAAGAGTTACTTTTCTCCCTCCATGAAAGTGTTTTTTAAGACTGACAAGGCCGG 3369
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219418 GAATGAAAAGAGTTACTTTTCTCCCTCCATGAAAGTGTTTTTTAAGACTGACAAGGCCGG 219477

Qy      3370 TGGTGGAGTGAGTGTTGTTGTTTTGTTTCGCAATTTTCTTACCAAGTGAAATTACTGACC 3429
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219478 TGGTGGAGTGAGTGTTGTTGTTTTGTTTCGCAATTTTCTTACCAAGTGAAATTACTGACC 219537

Qy      3430 ATAATTTTATTTAATAGTTTTTAGTTTTTGAAAGCCTGTGTTTTTCTTTGCTTTTAAACTA 3489
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219538 ATAATTTTATTTAATAGTTTTTAGTTTTTGAAAGCCTGTGTTTTTCTTTGCTTTTAAACTA 219597

Qy      3490 CAGGCGACCACTCTGGGATAGGATTCCACTCCCAAAAGTAATGTGTTGGGTGTGGTGGCA 3549
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219598 CAGGCGACCACTCTGGGATAGGATTCCACTCCCAAAAGTAATGTGTTGGGTGTGGTGGCA 219657

Qy      3550 CACGCCTGTGATCTCAGGACTCAGGAGGTGAAGCCAGGGGAACCCCAAGTTCCCGGTCAG 3609
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    219658 CACGCCTGTGATCTCAGGACTCAGGAGGTGAAGCCAGGGGAACCCCAAGTTCCCGGTCAG 219717

Qy      3610 CCTAGGCTACATAGTGAGACTTTGTCTTAACAAACAGAAAAAAAGAAAAAGAAAGAAAAA 3669
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219718 CCTAGGCTACATAGTGAGACTTTGTCTTAACAAACAGAAAAAAAGAAAAAGAAAGAAAAA 219777

Qy      3670 AAAAAGAAAGAAAAAAAAAARAAAGAAAAATGGAGCAGTATGAGTTATGGGTGCCAGTTTC 3729
             ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db    219778 AAAAAGAAAGAAAAAAAAAGAAAGAAAAATGGAGCAGTATGAGTTATGGGTGCCAGTTTC 219837

Qy      3730 TGCAGGAGACTTTGGGAGTAAGAGCTGCCTCTGTTCCTAAGAGGCTGCAAGGCTGCGGCT 3789
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219838 TGCAGGAGACTTTGGGAGTAAGAGCTGCCTCTGTTCCTAAGAGGCTGCAAGGCTGCGGCT 219897

Qy      3790 GGGGGGCCAGTTCAGGTGCCAGGTGAGGTACGGCCACATCTATTTGCCTGGCTCTTTGAC 3849
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219898 GGGGGGCCAGTTCAGGTGCCAGGTGAGGTACGGCCACATCTATTTGCCTGGCTCTTTGAC 219957

Qy      3850 AAAGAGAGCAGCTTAGAGTGGCTAGAGTGTGATGAGTTTTACACAATGGCGATGTGTGTG 3909
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219958 AAAGAGAGCAGCTTAGAGTGGCTAGAGTGTGATGAGTTTTACACAATGGCGATGTGTGTG 220017

Qy      3910 TGTGCCCCTGGCACAGCTCAATAAAGTTTCTGTCTCACTTTA 3951
             |||||||||||||||||||||||||||||||||||||||||||
Db    220018 TGTGCCCCTGGCACAGCTCAATAAAGTTTCTGTCTCACTTTA 220059
```

```
RESULT 9
AF336369
LOCUS       AF336369               3273 bp    mRNA    linear   VRT 02-JAN-2002
DEFINITION  Danio rerio piwi protein mRNA, complete cds.
ACCESSION   AF336369
VERSION     AF336369.1  GI:18033486
KEYWORDS    .
SOURCE      Danio rerio.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 3273)
  AUTHORS   Weeraratne,S.D., Gong,Z. and Tan,C.-H.
  TITLE     Cloning and characterization of zebrafish homolog of piwi,
            essential for germ-line stem cell self-renewal
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3273)
  AUTHORS   Weeraratne,S.D., Gong,Z. and Tan,C.-H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2001) Department of Biological Sciences, National
            University of Singapore, 10 Kent Ridge Crescent, Singapore 119260,
            Singapore
FEATURES             Location/Qualifiers
     source          1. .3273
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
     CDS             64. .2640
                     /codon_start=1
                     /product="piwi protein"
                     /protein_id="AAL57170.1"
                     /db_xref="GI:18033487"
                     /translation="MTGRARARSRGRGRGQEPAAPGAQPPVSQEAAKPVVSTPSEGQL
                     VGRGRQKPAPGAMSEEAMLQISAGFQQVKIGERGGRRRDFHDSGVHTRQLMEHVKESK
                     TGVSGTAIELRANFMRLLSRPMWALYQYHVDYKPPMESRRLRSALLFQHEETLGKAHT
                     FDGAILFLPNKLRNAETVLCSETRNGEKVEITVTLTNELPPSSPVCLQFYNILFRRIL
                     RILNMQQIGRHYYNPDDPFNIPQHRLTIWPGFMTTILQYESSIMLCSDVSHKVLRSET
                     VLDFMYSLRQQCGDQRFPEACTKELVGLIILTKYNNKTYRIDDIAWDHTPNNTFKKGD
                     TEISFKNYFKSQYGLDITDGNQVLLVSHVKRLGPSGRPPPGPAMLVPEFCYLTGLTDK
                     MRADFNIMKDLASHTRLSPEQREGRINRLISNINRNGDVQNELTTWGLSFENKLLSLN
                     GRVLPSERIIQQGRAFEYNPWTADWSKEMRGLPLISCMSLDNWLMFYTRRNADVAQSL
                     LQTLNKVSGPMGIRMQRAVMIEYEDRQESLLRALQQNVARETQMVVVILPTNRKDKYD
                     CVKKYLCVDCPTPSQCVVSRTISKPQALMTVATKIALQMNCKMGGELWSVEIPLRQLM
                     IVGIDCYHDTAAGKRSIGAMVASLNQGMSRWFSKCVLQNRGQEIIDALKGSLQGALKA
                     YLKYNNSLPSRIIVYRDGVGDGMLQSVVDYEVPQIMQSIKTMGQDYEPKLSVVVVKKR
                     ISSRFFARIDGKIANPPPGTVIDTEVTRPEWYDFFIVSQAVRFGCVAPTHYNVVFDNS
```

GLKPDHMQRLTYKLCHMYYNWQGIVRVPAPCQYAHKLAFLVGQSIHKEPNMNLDDFLY
YL"

BASE COUNT 922 a 717 c 787 g 847 t
ORIGIN

Query Match 27.5%; Score 1119.2; DB 5; Length 3273;
Best Local Similarity 66.6%; Pred. No. 3.3e-194;
Matches 1610; Conservative 0; Mismatches 806; Indels 3; Gaps 1;

```
Qy   362 GGAGCCACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACTG 421
Db   226 GGAGCCATGTCTGAAGAAGCGATGCTCCAGATTTCTGCTGGCTTTCAGCAGGTGAAGATC 285

Qy   422 GCAGAGAGAGGAGGGCGTCGCCGAGACTTCCATGACNTTGGTGTGAACACCAGACAGAAC 481
Db   286 GGGGAGAGAGGTGGACGCAGACGGGATTTCCACGATTCTGGTGTCCATACTCGTCAACTT 345

Qy   482 CTTGACCATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACC 541
Db   346 ATGGAACACGTGAAAGAATCCAAAACAGGTGTTTCTGGGACTGCCATTGAGCTGAGGGCC 405

Qy   542 AACCACTTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTAC 601
Db   406 AATTTTATGCGTCTGCTTTCACGTCCTATGTGGGCTTTGTATCAGTATCATGTGGACTAC 465

Qy   602 AATCCCCTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTC 661
Db   466 AAGCCACCAATGGAGTCCAGACGACTGCGTTCGGCTTTGCTTTTCCAGCACGAGGAGACT 525

Qy   662 ATTGGAAGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCAC 721
Db   526 CTGGGGAAAGCACACACCTTTGATGGAGCCATTCTCTTTCTCCCCAACAAACTGCGCAAC 585

Qy   722 AAGGTCACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCACC 781
Db   586 GCTGAGACAGTCCTGTGCAGTGAGACCAGAAATGGAGAGAAGGTTGAGATCACTGTGACA 645

Qy   782 CTGACCAACGAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTTC 841
Db   646 CTGACCAATGAGCTGCCGCCGTCCTCTCCTGTATGCCTGCAGTTTTACAACATCCTCTTC 705

Qy   842 AGGAGGCTCTTGAAAATCATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAGT 901
Db   706 AGGAGAATCCTAAGAATACTAAACATGCAGCAGATCGGACGTCATTACTACAATCCTGAT 765

Qy   902 GACCCGATTGATATTCCAAACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCATC 961
Db   766 GACCCATTCAACATCCCACAGCACAGGCTGACAATTTGGCCTGGCTTCATGACCACCATC 825

Qy   962 CTTCAGTATGAGAACAACATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAGC 1021
Db   826 CTACAGTACGAGTCCAGCATCATGCTGTGCTCTGATGTGAGCCACAAGGTTCTGCGCAGT 885

Qy  1022 GAGACTGTCCTAGACTTCATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCAG 1081
Db   886 GAGACGGTTCTTGACTTTATGTACAGTCTGAGACAGCAGTGTGGAGACCAGCGCTTCCCT 945

Qy  1082 GAGCAAGTGTCGAAGGAGCTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGACC 1141
Db   946 GAAGCCTGCACGAAGGAGCTTGTGGGCCTGATCATTCTGACCAAGTACAACAACAAGACA 1005

Qy  1142 TACCGGGTGGATGACATTGACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGAT 1201
Db  1006 TACAGGATTGATGACATTGCCTGGGATCATACTCCCAACAACACATTCAAGAAAGGAGAT 1065

Qy  1202 GGCTCGGAGGTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGAC 1261
Db  1066 ---ACAGAAATTTCATTCAAGAACTACTTCAAATCGCAATATGGTCTTGACATAACAGAT 1122

Qy  1262 CTGAAGCAGCCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCTG 1321
Db  1123 GGCAACCAGGTTCTTCTCGTTAGCCATGTGAAGAGACTTGGCCCCTCAGGACGCCCTCCA 1182

Qy  1322 CCTGGCCCAGCTATGNTCATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAATG 1381
Db  1183 CCGGGACCAGCTATGCTTGTCCCCGAGTTCTGCTACCTCACTGGTCTAACCGATAAAATG 1242

Qy  1382 CGCAATGATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCAG 1441
Db  1243 CGTGCTGATTTTAACATCATGAAGGATCTTGCTTCACATACACGGCTAAGTCCTGAACAG 1302

Qy  1442 CGGCAGCGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAGA 1501
Db  1303 AGAGAGGGCAGAATTAACCGCTTAATCTCCAACATCAACAGAAATGGAGACGTGCAGAAT 1362

Qy  1502 GAGCTTCGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAATC 1561
Db  1363 GAACTGACTACTTGGGGTCTGAGCTTTGAGAACAAGCTTCTTAGTCTTAATGGGAGAGTT 1422

Qy  1562 TTACAATCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGCA 1621
Db  1423 TTGCCCTCTGAGAGGATCATACAGGGAGGCAGAGCATTTGAGTATAACCCATGGACAGCA 1482

Qy  1622 GACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGG 1681
Db  1483 GACTGGTCTAAAGAGATGAGAGGTCTCCCTCTGATTAGCTGCATGTCTCTGGACAACTGG 1542

Qy  1682 CTGCTGATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTTC 1741
Db  1543 CTCATGTTCTACACGCGCAGAAATGCAGATGTTGCCCAATCTCTGCTTCAGACTCTCAAC 1602

Qy  1742 AAAGTGACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGAC 1801
Db  1603 AAAGTGTCTGGACCCATGGGGATCCGCATGCAGAGGGCTGTCATGATTGAGTATGAAGAT 1662

Qy  1802 AGAACAGAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTT 1861
Db  1663 CGTCAGGAATCTCTCCTTAGAGCTTTGCAACAAAACGTAGCACGTGAAACTCAGATGGTG 1722

Qy  1862 GTCTGTCTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTGT 1921
Db  1723 GTGGTGATCTTGCCTACCAATCGAAAAGACAAGTACGACTGTGTAAAGAAGTACCTGTGT 1782

Qy  1922 ACAGACTGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACA 1981
Db  1783 GTGGACTGCCCTACTCCTAGTCAGTGTGTGGTGTCTCGCACCATCAGCAAACCTCAAGCA 1842

Qy  1982 GTCATGGCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTC 2041
Db  1843 CTTATGACTGTGGCCACCAAGATTGCTTTGCAGATGAACTGCAAAATGGGAGGAGAGCTG 1902

Qy  2042 TGGCGGGTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGAC 2101
Db  1903 TGGAGTGTAGAAATTCCACTCCGGCAGCTCATGATTGTAGGTATAGACTGCTACCATGAT 1962

Qy  2102 ACCACAGCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGACC 2161
Db  1963 ACCGCTGCTGGAAAAAGGTCCATCGGCGCAATGGTGGCCAGTCTAAATCAGGGCATGTCA 2022

Qy  2162 CGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAG 2221
Db  2023 AGGTGGTTCTCCAAGTGTGTCTTGCAAAACCGTGGACAGGAGATCATCGATGCACTGAAG 2082

Qy  2222 GTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGT 2281
Db  2083 GGGTCATTGCAGGGTGCTCTGAAGGCCTATCTGAAATACAACAATTCTCTGCCCTCGCGC 2142

Qy  2282 GTCATCGTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGAG 2341
Db  2143 ATCATTGTTTACCGAGATGGAGTTGGAGATGGCATGCTGCAGAGCGTTGTGGACTATGAA 2202

Qy  2342 GTCCCACAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGACT 2401
Db  2203 GTTCCTCAGATCATGCAGTCTATCAAGACCATGGGGCAAGATTATGAGCCCAAGCTCTCA 2262

Qy  2402 GTAATCGTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTT 2461
```

```
Db      2263 GTAGTGGTGGTGAAAAAGCGCATCAGCTCCCGATTCTTTGCTCGGATTGATGGCAAAATT 2322

Qy      2462 CAGAACCCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGAC 2521

Db      2323 GCCAATCCCCCTCCTGGGACAGTCATTGACACTGAAGTCACTCGTCCAGAGTGGTATGAT 2382

Qy      2522 TTTTTCATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGTC 2581

Db      2383 TTCTTCATTGTAAGCCAGGCTGTGCGCTTTGGGTGTGTCGCACCTACTCACTACAATGTG 2442

Qy      2582 ATCTATGACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCTGC 2641

Db      2443 GTGTTTGACAACAGCGGCCTTAAACCAGATCACATGCAGAGACTCACCTACAAACTTTGC 2502

Qy      2642 CACGTGTACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACAC 2701

Db      2503 CACATGTACTACAACTGGCAGGGGATTGTGCGAGTGCCAGCTCCATGCCAGTATGCCCAC 2562

Qy      2702 AAGCTGGCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAAC 2761

Db      2563 AAATTGGCCTTCCTGGTTGGCCAGAGCATTCACAAGGAGCCTAACATGAATCTGGATGAC 2622

Qy      2762 CGCCTCTACTACCTCTAAC 2780

Db      2623 TTCCTGTATTACCTGTAAC 2641
```

```
RESULT 10
BC031060
LOCUS       BC031060                3006 bp    mRNA    linear   PRI 13-JUN-2002
DEFINITION  Homo sapiens, Similar to piwi like homolog 1 (Drosophila), clone
            MGC:33296 IMAGE:5271278, mRNA, complete cds.
ACCESSION   BC031060
VERSION     BC031060.1  GI:21410560
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3006)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 48 Row: b Column: 15
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES              Location/Qualifiers
     source           1. .3006
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="MGC:33296 IMAGE:5271278"
                      /tissue_type="Testis"
                      /clone_lib="NIH_MGC_97"
                      /lab_host="DH10B"
                      /note="Vector: pBluescript"
     CDS              4. .2562
                      /codon_start=1
                      /product="Similar to piwi like homolog 1 (Drosophila)"
                      /protein_id="AAH31060.1"
                      /db_xref="GI:21410561"
                      /translation="MSGRARVKARGIARSPSATEVGRIQASPLPRSVDLSNNEASSSN
                      GFLGTSRISTNDKYGISSGDAGSTFMERGVKNKQDFMDLSICTREKLAHVRNCKTGSS
                      GIPVKLVTNLFNLDFPQDWQLYQYHVTYIPDLASRRLRIALLYSHSELSNKAKAFDGA
                      ILFLSQKLEEKVTELSSETQRGETIKMTITLKRELPSSSPVCIQVFNIIFRKILKKLS
                      MYQIGRNFYNPSEPMEIPQHKLSLWPGFAISVSYFERKLLFSADVSYKVLRNETVLEF
                      MTALCQRTGLSCFTQTCEKQLIGLIVLTRYNNRTYSIDDIDWSVKPTHTFQKRDGTEI
                      TYVDYYKQQYDITVSDLNQPMLVSLLKKKRNDNSEAQLAHLIPELCFLTGLTDQATSD
                      FQLMKAVAEKTRLSPSGRQQRLARLVDNIQRNTNARFELETWGLHFGSQISLTGRIVP
                      SEKILMQDHICQPVSAADWSKDIRTCKILNAQSLNTWLILCSDRTEYVAESFLNCLRR
                      VAGSMGFNVDYPKIIKVQENPAAFVRAIQQYVDPDVQLVMCILPSNQKTYYDSIKKYL
                      SSDCPVPSQCVLARTLNKQGMMMSIATKIAMQMTCKLGGELWAVEIPLKSLMVVGIDV
                      CKDALSKDVMVVGCVASVNPRITRWFSRCILQRTMTDVADCLKVFMTGALNKWYKYNH
                      DLPARIIVYRAGVGDGQLKTLIEYEVPQLLSSVAESSSNTSSRLSVIVVRKKCMPRFF
                      TEMNRTVQNPPLGTVVDSEATRNEWYDFYLISQVACRGTVSPTYYNVIYDDNGLKPDH
                      MQRLTFKLCHLYYNWPGIVSVPAPCQYAHKLTFLVAQSIHKEPSLELANHLFYL"
BASE COUNT      904 a    627 c    674 g    801 t
ORIGIN
```

```
  Query Match             18.1%;  Score 734.8;  DB 9;  Length 3006;
  Best Local Similarity   58.0%;  Pred. No. 4e-124;
  Matches 1354;  Conservative    0;  Mismatches  971;  Indels    9;  Gaps    3;

Qy       446 GACTTCCATGACNTTGGTGTGAACACCAGACAGAACCTTGACCATGTCAAAGAGTCAAAG 505

Db       238 GACTTTATGGATTTGAGTATCTGTACCAGAGAAAAATTGGCACATGTGAGAAATTGTAAA 297

Qy       506 ACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCACTTCCGGCTGACCTCGCGC 565

Db       298 ACAGGTTCCAGTGGAATACCTGTGAAACTGGTTACAAACCTCTTTAACTTAGATTTTCCC 357

Qy       566 CCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCCCTGATGGAGGCCCGAAGG 625

Db       358 CAAGACTGGCAGCTATACCAGTACCATGTGACATATATTCCAGATTTAGCATCTAGAAGG 417

Qy       626 CTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGAAGGTGTCATGCTTTCGAT 685

Db       418 CTGAGAATTGCTTTACTTTATAGTCATAGTGAACTTTCCAACAAAGCAAAAGCATTCGAC 477

Qy       686 GGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGTCACAGAAGTATTCAGTCAG 745

Db       478 GGTGCCATCCTTTTTCTGTCACAAAAGCTAGAAGAAAAGGTCACAGAGTTGTCAAGTGAA 537

Qy       746 ACTCGGAATGGGGAACACGTGAGGATCACCATCACCCTGACCAACGAGCTGCCGCCCACC 805

Db       538 ACTCAAAGAGGTGAGACTATAAAGATGACTATCACCCTGAAGAGGGAGCTGCCATCAAGT 597

Qy       806 TCGCCCACCTGCCTGCAGTTCTATAACATCNTCTTCAGGAGGCTCTTGAAAATCATGAAT 865

Db       598 TCTCCCGTGTGCATCCAGGTCTTCAATATCATCTTCAGAAAGATCCTCAAAAAGTTGTCC 657

Qy       866 TTGCAACAAATTGGACGGAATTATTACAATCCAAGTGACCCGATTGATATTCCAAACCAC 925

Db       658 ATGTACCAAATTGGACGGAACTTCTATAATCCTTCAGAGCCAATGGAAATTCCCCAGCAC 717

Qy       926 AGGTTGGTGATCTGGCCCGGCTTCACCACCTCCATCCTTCAGTATGAGAACAACATCATG 985

Db       718 AAATTATCCCTTTGGCCTGGGTTTGCCATTTCTGTGTCATATTTTGAAAGGAAGCTCCTG 777

Qy       986 CTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAGCGAGACTGTCCTAGACTTCATGTTC 1045

Db       778 TTTAGTGCTGATGTGAGTTACAAAGTCCTCCGGAATGAGACGGTTCTGGAATTCATGACT 837

Qy      1046 AATCTATACCAGCAGACAGAGGAGCACAAGTTCCAGGAGCAAGTGTCGAAGGAGCTCATA 1105

Db       838 GCTCTCTGTCAAAGAACTGGCTTGTCCTGTTTCACCCAGACGTGTGAGAAGCAGCTAATA 897
```

```
Qy      1106 GGCCTCATCGTTCTCACCAAGTACAATAACAAGACCTACCGGGTGGATGACATTGACTGG 1165
             || ||||| || || || |  |||||||||| ||||||    | ||||||||||||||||
Db       898 GGGCTCATTGTCCTTACAAGATACAATAACAGAACCTACTCCATTGATGACATTGACTGG 957

Qy      1166 GACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGATGGCTCGGAGGTCAGCTTCCTGGAG 1225
               ||| || |    |||||| ||||| ||||||| | ||| ||| ||   ||||
Db       958 TCAGTGAAGCCCACACACACCTTTCAGAAGCGGGATGGCACCGAGATCACCTATGTGGAT 1017

Qy      1226 TACTACAGGAAGCAATACAACCAGGAGATCACGGACCTGAAGCAGCCGGTGCTGGTGAGC 1285
             ||||||| | |||| ||  |         |  |||||||||| |||||  |||| || ||
Db      1018 TACTACAAGCAGCAGTATGATATTACTGTATCGGACCTGAATCAGCCCATGCTTGTTAGT 1077

Qy      1286 CAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCTGCCTGGCCCAGCTATGNTCATCCCT 1345
             |     |||  || |||       | |  ||    | ||    |  ||      | || |||
Db      1078 CTGTTAAAGAAGAAGAGA---AATGACAACAGTGAGGCTCAGCTCGCCCACCTGATACCT 1134

Qy      1346 GAACTCTGCTATCTCACAGGCCTGACTGATAAAATGCGCAATGATTTCAATGTGATGAAG 1405
             || ||||||| ||| ||||| |||||||| |          |||||| | ||||||||
Db      1135 GAGCTCTGCTTTCTAACAGGGCTGACTGACCAGGCAACATCTGATTTCCAGCTGATGAAG 1194

Qy      1406 GACCTGGCAGTGCACACGCGGCTGACCCCTGAGCAGCGGCAGCGGGAGGTGGGCCGCCTC 1465
             |  |||| |   | || || || |  |||      |||||| |    ||| | | ||
Db      1195 GCTGTGGCTGAAAAGACACGTCTCAGTCCTTCAGGCCGGCAGCAGCGCCTGGCCAGGCTT 1254

Qy      1466 ATCGACTACATCCACAAGGATGACAATGTGCAGAGAGAGCTTCGAGACTGGGGCCTGAGC 1525
             | ||| |||||| | | ||  |||||  |     || ||      |||||| |||
Db      1255 GTGGACAACATCCAGAGGAATACCAATGCTCGCTTTGAACTAGAGACCTGGGGACTGCAT 1314

Qy      1526 TTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAATCTTACAATCTGAGAAGATCCACCAG 1585
             || |   || |  | | ||  |  |||| | ||  | |  || || || ||       |
Db      1315 TTTG---GAAGCCAGATATCTCTGACTGGCCGGATTGTGCCTTCAGAAAAAATATTAATG 1371

Qy      1586 GGCGGAAAGACGTTTGATTACAACCCACAATTTGCAGACTGGTCCAAAGAAACAAGAGGC 1645
               |  | |  | | |          |      ||| |||||||||| || |   ||
Db      1372 CAAGACCACATATGTCAACCTGTGTC---TGCTGCTGACTGGTCCAAGGATATTCGAACT 1428

Qy      1646 GCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGGCTGCTGATCTATACCCGCAGGAAT 1705
               | |  | |  |   || | ||| ||| |||| || |  | | || |  ||| | |
Db      1429 TGCAAGATTTTAAATGCACAGTCTTTGAATACCTGGTTGATTTTATGTAGCGACAGAACT 1488

Qy      1706 TATGAAGCAGCCAACTCACTGATACAGAACCTGTTCAAAGTGACTCCAGCCATGGGCATC 1765
               |  | |  ||| |     |  |  |   | ||   | |||  |     |||||||  |
Db      1489 GAATATGTTGCCGAGAGCTTTCTGAACTGCTTGAGAAGAGTTGCAGGTTCCATGGGATTT 1548

Qy      1766 CAGATGAAAAAGGCAATCATGATCGAGGTGGATGACAGAACAGAAGCTTATCTGAGAGCC 1825
               |  || |  | |  || || | || |  |  |||  || | | |  |||||
Db      1549 AATGTGGACTACCCCAAAATCATAAAAGTACAAGAAAATCCAGCTGCATTTGTTAGAGCT 1608

Qy      1826 TTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGTCTCTTGTCAAGTAATCGG 1885
               | |||||  | ||     | ||   |||| | ||  | ||  |  || |   ||||| |
Db      1609 ATACAGCAATATGTTGATCCTGATGTTCAGCTGGTAATGTGCATTCTGCCTTCTAATCAG 1668

Qy      1886 AAGGACAAATATGATGCCATCAAGAAGTACTTGTGTACAGACTGCCCCACCCCAAGTCAG 1945
             |||  | | |||||| |||| || || || ||| |  ||||||||||   |||||| ||
Db      1669 AAGACCTATTATGATTCCATTAAAAAATATTTGAGCTCAGACTGCCCAGTCCCAAGCCAA 1728

Qy      1946 TGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATGGCCATTGCCACCAAGATC 2005
             |||||| | || |||||| ||   || |||   |  | |||   || ||||||||||||
Db      1729 TGTGTGCTTGCTCGGACCTTGAATAAACAGGGCATGATGATGAGTATCGCCACCAAGATC 1788

Qy      2006 GCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGGGTGGACATGGCCCTGAAA 2065
             ||  |||||||||  |||||| | |||||||||||| |||   ||||| ||  |  | ||
Db      1789 GCTATGCAGATGACTTGCAAGCTCGGAGGCGAGCTGTGGGCTGTGGAAATACCTTTAAAG 1848

Qy      2066 CTGGCAATGATCGTGGGCATCGACTGTTACCATGACACCACAGCTGGGCGGAGGTCCATC 2125
               ||| | || || || ||     |  | ||  |           |     |
Db      1849 TCCCTGATGGTGGTCGGTATTGATGTCTGTAAAGATGCACTCAGCAAGGACGTGATGGTT 1908

Qy      2126 GCAGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCCCGCTGCGTCTTT 2185
               |  |||| ||| ||  | ||     | || ||| | ||||| ||||||||  || ||
Db      1909 GTTGGATGCGTGGCAAGTGTTAACCCCAGAATCACCAGGTGGTTTTCCCGCTGTATCCTT 1968

Qy      2186 CAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGCTTGCAAGCTGCTCTGAGG 2245
             |||          |   ||  | |  ||| |  | || || | | ||   |  || || |
Db      1969 CAGAGAACAATGACTGATGTTGCAGATTGCTTGAAAGTTTTCATGACTGGAGCACTCAAC 2028

Qy      2246 GCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATCGTGTACCGAGACGGTGTG 2305
               |||       | |||| |  |  ||||    ||  | || |||||||| |  |||||
Db      2029 AAATGGTACAAGTACAATCATGATTTGCCAGCACGGATAATTGTGTACCGTGCTGGTGTA 2088

Qy      2306 GGGGACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCACAGTTCCTAGATTGCCTC 2365
             ||||| || |||||||| || ||  |  | |||||  |||||||||  | ||    |  |
Db      2089 GGGGATGGTCAGCTGAAAACACTTATTGAATATGAAGTCCCACAGCTGCTGAGCAGTGTG 2148

Qy      2366 AAGTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATCGTGGTGAAGAAGCGTGTC 2425
                 |  | |   |  | || | |||||||| | || || ||||| | ||||     |
Db      2149 GCAGAATCCAGCTCAAATACCAGCTCAAGACTGTCGGTGATTGTGGTCAGGAAGAAGTGC 2208

Qy      2426 AATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAACCCTCTTCCAGGGACAGTC 2485
             |   |  | || ||| |  |         | |   | |||||||| |  |  || || ||
Db      2209 ATGCCACGATTCTTTACCGAAATGAACCGCACTGTACAGAACCCCCCACTTGGCACTGTT 2268

Qy      2486 ATCGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTCATCGTGAGCCAGGCAGTG 2545
               | |||   ||||  ||  |    || |||||||||||||   |  | |||||||  |
Db      2269 GTGGATTCAGAAGCAACACGTAACGAATGGTATGACTTTTATCTGATCAGCCAGGTGGCC 2328

Qy      2546 AGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTATGACAGCAGTGGCCTGAAG 2605
               |  | || | |||     || ||  |||| ||||||||||||||||   ||  ||| |||||
Db      2329 TGCCGGGGAACTGTTAGTCCTACCTACTATAATGTCATCTATGATGACAACGGCTTGAAG 2388

Qy      2606 CCCGACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTGTACTATAATTGGCCTGGA 2665
             |||||||| || ||| | || |||| ||| .| ||||||  ||||||| || ||||| ||
Db      2389 CCCGACCATATGCAGAGACTTACATTCAAATTGTGCCACCTGTACTACAACTGGCCGGGC 2448

Qy      2666 GTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTGGCCTTCCTCGTGGGCCAG 2725
               |  || | ||||| ||||| || |||||||| |||||||| |||| || ||||  ||
Db      2449 ATAGTCAGTGTCCCAGCACCATGTCAGTATGCTCACAAGCTGACCTTTCTGGTGGCACAA 2508

Qy      2726 AGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTCTACTACCTCTAA 2779
             ||||| || | ||| || |  ||     |  ||||||  |||| |||||| | |
Db      2509 AGCATTCATAAAGAACCCAGTCTGGAATTAGCCAACCATCTCTTCTACCTGTGA 2562
```

```
RESULT 11
AK093475
LOCUS       AK093475               3138 bp    mRNA    linear   PRI 15-JUL-2002
DEFINITION  Homo sapiens cDNA FLJ36156 fis, clone TESTI2025546, weakly similar
            to Mus musculus mRNA for MIWI (piwi).
ACCESSION   AK093475
VERSION     AK093475.1  GI:21752360
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens testis cDNA to mRNA, clone_lib:TESTI2
            clone:TESTI2025546.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S.,
            Ono,Y., Hotuta,T., Hiraoka,S., Murakawa,K., Takiguchi,S.,
            Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M.,
            Yamashita,H., Chiba,Y., Sugiyama,T., Irie,R., Otsuki,T., Sato,H.,
            Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y.,
            Saito,K., Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y.,
            Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3138)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
```

Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

```
FEATURES             Location/Qualifiers
     source          1. .3138
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="TESTI2025546"
                     /tissue_type="testis"
                     /clone_lib="TESTI2"
                     /note="cloning vector: pME18SFL3"
     CDS             152. .2710
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAC04179.1"
                     /db_xref="GI:21752361"
                     /translation="MSGRARVKARGIARSPSATEVGRIQASPLPRSVDLSNNEASSSN
                     GFLGTSRISTNDKYGISSGDAGSTFMERGVKNKQDFMDLSICTREKLAHVRNCKTGSS
                     GIPVKLVTNLFNLDFPQDWQLYQYHVTYIPDLASRRLRIALLYSHSELSNEAKAFDGA
                     ILFLSQKLEEKVTELSSETQRGETIKMTITLKRELPSSSPVCIQVFNIIFRKILKKLS
                     MYQIGRNFYNPSEPMEIPQHKLSLWPGFAISVSYFERKLLFSADVSYKVLRNETVLEF
                     MTALCQRTGLSCFTQTCEKQLIGLIVLTRYNNRTYSIDDIDWSVKPTHTFQKRDGTEI
                     TYVDYYKQQYDITVSDLNQPMLVSLLKKKRNDNSEAQLAHLIPELCFLTGLTDQATSD
                     FQLMKAVAEKTRLSPSGRQQRLARLVDNIQRNTNARFELETWGLHFGSQISLTGRIVP
                     SEKILMQDHICQPVSAADWSKDIRTCKILNAQSLNTWLILCSDRTEYVAESFLNCLRR
                     VAGSMGFNVDYPKIIKVQENPAAFVRAIQQYVDPDVQLVMCILPSNQKTYYDSIKKYL
                     SSDCPVPSQCVLARTLNKQGMMMSIATKIAMQMTCKLGGELWAVEIPLKSLMVVGIDV
                     CKDALSKDVMVVGCVASVNPRITRWFSRCILQRTMTDVADCLKVFMTGALNKWYKYNH
                     DLPARIIVYRAGVGDGQLETLIEYEVPQLLSSVAESGSNTSSRLSVIVVRKKCMPRFF
                     TEMNRTVQNPPLGTVVDSEATRNEWYDFYLISQVACRGTVSPTYYNVIYDDNGLKPDH
                     MQRLTFKLCHLYYNWPGIVSVPAPCQYAHKLTFLVAQSIHKEPSLELANHLFYL"
BASE COUNT       912 a    678 c    714 g    834 t
ORIGIN
```

Query Match 18.1%; Score 734.8; DB 9; Length 3138;
Best Local Similarity 58.0%; Pred. No. 4e-124;
Matches 1354; Conservative 0; Mismatches 971; Indels 9; Gaps 3;

```
Qy        446 GACTTCCATGACNTTGGTGTGAACACCAGACAGAACCTTGACCATGTCAAAGAGTCAAAG 505
Db        386 GACTTTATGGATTTGAGTATCTGTACCAGAGAAAAATTGGCACATGTGAGAAATTGTAAA 445

Qy        506 ACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCACTTCCGGCTGACCTCGCGC 565
Db        446 ACAGGTTCCAGTGGAATACCTGTGAAACTGGTTACAAACCTCTTTAACTTAGATTTTCCC 505

Qy        566 CCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCCCTGATGGAGGCCCGAAGG 625
Db        506 CAAGACTGGCAGCTATACCAGTACCATGTGACATATATTCCAGATTTAGCATCTAGAAGG 565

Qy        626 CTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGAAGGTGTCATGCTTTCGAT 685
Db        566 CTGAGAATTGCTTTACTTTATAGTCATAGTGAACTTTCCAACGAAGCAAAAGCATTCGAC 625

Qy        686 GGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGTCACAGAAGTATTCAGTCAG 745
Db        626 GGTGCCATCCTTTTTCTGTCACAAAAGCTAGAAGAAAAGGTCACAGAGTTGTCAAGTGAA 685

Qy        746 ACTCGGAATGGGGAACACGTGAGGATCACCATCACCCTGACCAACGAGCTGCCGCCCACC 805
Db        686 ACTCAAAGAGGTGAGACTATAAAGATGACTATCACCCTGAAGAGGGAGCTGCCATCAAGT 745

Qy        806 TCGCCCACCTGCCTGCAGTTCTATAACATCNTCTTCAGGAGGCTCTTGAAAATCATGAAT 865
Db        746 TCTCCCGTGTGCATCCAGGTCTTCAATATCATCTTCAGAAAGATCCTCAAAAAGTTGTCC 805

Qy        866 TTGCAACAAATTGGACGGAATTATTACAATCCAAGTGACCCGATTGATATTCCAAACCAC 925
Db        806 ATGTACCAAATTGGACGGAACTTCTATAATCCTTCAGAGCCAATGGAAATTCCCCAGCAC 865

Qy        926 AGGTTGGTGATCTGGCCCGGCTTCACCACCTCCATCCTTCAGTATGAGAACAACATCATG 985
Db        866 AAATTATCCCTTTGGCCTGGGTTTGCCATTTCTGTGTCATATTTTGAAAGGAAGCTCCTG 925

Qy        986 CTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAGCGAGACTGTCCTAGACTTCATGTTC 1045
Db        926 TTTAGTGCTGATGTGAGTTACAAAGTCCTCCGGAATGAGACGGTTCTGGAATTCATGACT 985

Qy       1046 AATCTATACCAGCAGACAGAGGAGCACAAGTTCCAGGAGCAAGTGTCGAAGGAGCTCATA 1105
Db        986 GCTCTCTGTCAAAGAACTGGCTTGTCCTGTTTCACCCAGACGTGTGAGAAGCAGCTAATA 1045

Qy       1106 GGCCTCATCGTTCTCACCAAGTACAATAACAAGACCTACCGGGTGGATGACATTGACTGG 1165
Db       1046 GGGCTCATTGTCCTTACAAGATACAATAACAGAACCTACTCCATTGATGACATTGACTGG 1105

Qy       1166 GACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGATGGCTCGGAGGTCAGCTTCCTGGAG 1225
Db       1106 TCAGTGAAGCCCACACACACCTTTCAGAAGCGGGATGGCACCGAGATCACCTATGTGGAT 1165

Qy       1226 TACTACAGGAAGCAATACAACCAGGAGATCACGGACCTGAAGCAGCCGGTGCTGGTGAGC 1285
Db       1166 TACTACAAGCAGCAGTATGATATTACTGTATCGGACCTGAATCAGCCCATGCTTGTTAGT 1225

Qy       1286 CAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCTGCCTGGCCCAGCTATGNTCATCCCT 1345
Db       1226 CTGTTAAAGAAGAAGAGA---AATGACAACAGTGAGGCTCAGCTCGCCCACCTGATACCT 1282

Qy       1346 GAACTCTGCTATCTCACAGGCCTGACTGATAAAATGCGCAATGATTTCAATGTGATGAAG 1405
Db       1283 GAGCTCTGCTTTCTAACAGGGCTGACTGACCAGGCAACATCTGATTTCCAGCTGATGAAG 1342

Qy       1406 GACCTGGCAGTGCACACGCGGCTGACCCCTGAGCAGCGGCAGCGGGAGGTGGGCCGCCTC 1465
Db       1343 GCTGTGGCTGAAAAGACACGTCTCAGTCCTTCAGGCCGGCAGCAGCGCCTGGCCAGGCTT 1402

Qy       1466 ATCGACTACATCCACAAGGATGACAATGTGCAGAGAGAGCTTCGAGACTGGGGCCTGAGC 1525
Db       1403 GTGGACAACATCCAGAGGAATACCAATGCTCGCTTTGAACTAGAGACCTGGGGACTGCAT 1462

Qy       1526 TTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAATCTTACAATCTGAGAAGATCCACCAG 1585
Db       1463 TTTG---GAAGCCAGATATCTCTGACTGGCCGGATTGTGCCTTCAGAAAAAATATTAATG 1519

Qy       1586 GGCGGAAAGACGTTTGATTACAACCCACAATTTGCAGACTGGTCCAAAGAAACAAGAGGC 1645
Db       1520 CAAGACCACATATGTCAACCTGTGTC---TGCTGCTGACTGGTCCAAGGATATTCGAACT 1576

Qy       1646 GCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGGCTGCTGATCTATACCCGCAGGAAT 1705
Db       1577 TGCAAGATTTTAAATGCACAGTCTTTGAATACCTGGTTGATTTTATGTAGCGACAGAACT 1636

Qy       1706 TATGAAGCAGCCAACTCACTGATACAGAACCTGTTCAAAGTGACTCCAGCCATGGGCATC 1765
Db       1637 GAATATGTTGCCGAGAGCTTTCTGAACTGCTTGAGAAGAGTTGCAGGTTCCATGGGATTT 1696

Qy       1766 CAGATGAAAAAGGCAATCATGATCGAGGTGGATGACAGAACAGAAGCTTATCTGAGAGCC 1825
Db       1697 AATGTGGACTACCCCAAAATCATAAAAGTACAAGAAAATCCAGCTGCATTTGTTAGAGCT 1756

Qy       1826 TTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGTCTCTTGTCAAGTAATCGG 1885
Db       1757 ATACAGCAATATGTTGATCCTGATGTTCAGCTGGTAATGTGCATTCTGCCTTCTAATCAG 1816

Qy       1886 AAGGACAAATATGATGCCATCAAGAAGTACTTGTGTACAGACTGCCCCACCCCAAGTCAG 1945
Db       1817 AAGACCTATTATGATTCCATTAAAAAATATTTGAGCTCAGACTGCCCAGTCCCAAGCCAA 1876

Qy       1946 TGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATGGCCATTGCCACCAAGATC 2005
Db       1877 TGTGTGCTTGCTCGGACCTTGAATAAACAGGGCATGATGATGAGTATCGCCACCAAGATC 1936
```

```
Qy       2006 GCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGGGTGGACATGGCCCTGAAA 2065
Db       1937 GCTATGCAGATGACTTGCAAGCTCGGAGGCGAGCTGTGGGCTGTGGAAATACCTTTAAAG 1996

Qy       2066 CTGGCAATGATCGTGGGCATCGACTGTTACCATGACACCACAGCTGGGCGGAGGTCCATC 2125
Db       1997 TCCCTGATGGTGGTCGGTATTGATGTCTGTAAAGATGCACTCAGCAAGGACGTGATGGTT 2056

Qy       2126 GCAGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCCCGCTGCGTCTTT 2185
Db       2057 GTTGGATGCGTGGCCAGTGTTAACCCCAGAATCACCAGGTGGTTTTCCCGCTGTATCCTT 2116

Qy       2186 CAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGCTTGCAAGCTGCTCTGAGG 2245
Db       2117 CAGAGAACAATGACTGATGTTGCAGATTGCTTGAAAGTTTTCATGACTGGAGCACTCAAC 2176

Qy       2246 GCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATCGTGTACCGAGACGGTGTG 2305
Db       2177 AAATGGTACAAGTACAATCATGATTTGCCAGCACGGATAATTGTGTACCGTGCTGGTGTA 2236

Qy       2306 GGGGACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCACAGTTCCTAGATTGCCTC 2365
Db       2237 GGGGATGGTCAGCTGGAAACACTTATTGAATATGAAGTCCCACAGCTGCTGAGCAGTGTG 2296

Qy       2366 AAGTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATCGTGGTGAAGAAGCGTGTC 2425
Db       2297 GCAGAATCCGGCTCAAATACCAGCTCAAGACTGTCGGTGATTGTGGTCAGGAAGAAGTGC 2356

Qy       2426 AATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAACCCTCTTCCAGGGACAGTC 2485
Db       2357 ATGCCACGATTCTTTACCGAAATGAACCGCACTGTACAGAACCCCCCACTTGGCACTGTT 2416

Qy       2486 ATCGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTCATCGTGAGCCAGGCAGTG 2545
Db       2417 GTGGATTCAGAAGCAACACGTAACGAATGGTATGACTTTTATCTGATCAGCCAGGTGGCC 2476

Qy       2546 AGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTATGACAGCAGTGGCCTGAAG 2605
Db       2477 TGCCGGGGAACTGTTAGTCCTACCTACTATAATGTCATCTATGATGACAACGGCTTGAAG 2536

Qy       2606 CCCGACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTGTACTATAATTGGCCTGGA 2665
Db       2537 CCCGACCATATGCAGAGACTTACATTCAAATTGTGCCACCTGTACTACAACTGGCCGGGC 2596

Qy       2666 GTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTGGCCTTCCTCGTGGGCCAG 2725
Db       2597 ATAGTCAGTGTCCCAGCACCATGTCAGTATGCTCACAAGCTGACCTTTCTGGTGGCACAA 2656

Qy       2726 AGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTCTACTACCTCTAA 2779
Db       2657 AGCATTCATAAAGAACCCAGTCTGGAATTAGCCAACCATCTCTTCTACCTGTGA 2710
```

```
RESULT 12
AY014899
LOCUS       AY014899               3057 bp    mRNA    linear   INV 02-JAN-2001
DEFINITION  Strongylocentrotus purpuratus seawi mRNA, complete cds.
ACCESSION   AY014899
VERSION     AY014899.1  GI:12007639
KEYWORDS    .
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 3057)
  AUTHORS   Rodriguez,A.J. and Bonder,E.M.
  TITLE     Seawi - Cloning and Molecular Characterization of a Sea Urchin
            Homolog of Piwi
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3057)
  AUTHORS   Rodriguez,A.J. and Bonder,E.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-2000) Biological Sciences, Rutgers
            University-Newark, 101 Warren Street Room 135, Newark, NJ 07102,
            USA
FEATURES             Location/Qualifiers
     source          1. .3057
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
     CDS             181. .2745
                     /codon_start=1
                     /product="seawi"
                     /protein_id="AAG42533.1"
                     /db_xref="GI:12007640"
                     /translation="MDRRPGEQAPPPMQQAAPPQEASSGMPVPGRGRSRGGTTQPGPG
                     AVPKQEPRGELTSQIAKLAIGASGEGRVGGAVQDRRRRRDEIELEPTTRPENFVKQAI
                     AGDKIALIANGFKLKTKPDWQLYQYRVDFEPEILNPRARFALLKGHSALLGKGLTLDM
                     DTMYSLFKLAEKVTNLSAERKDGSNVNICVTHVATLNPMAPNTLHLYNVLFRRCLKMI
                     NMEQVGRNYYDPTAAIDIKQHGLQLWPGFVTSILQYEYDVMLLSDISHKVLRTQTVWE
                     VMNDLFNKARGRFKEEITKLMIGQIVLTKYNNKTYRVDDIDFDTTPADTFETRSGPVS
                     YVDYFKKSYERVIHDVNQPMLVSRPKKREEKKGVGPAYLPPELCFLTGLSDDMRADFN
                     VMKDMGQHTRVGPQDRCRTLSGFIKKLSSNEEVKTYLDSWGMEFDKEQVKLTGRVLPP
                     EKLFQRGKQFSYNPSNADWSRDTRGNALTDAKILNNWKIFYTRRDANRGQDFIKSLVR
                     VANPMGMNVRGPEIVELPDDRTETYTRSLQAQIAQDTQIVVVILPTNRKDRYDAIKKT
                     CVVTHPCPSQVIVSRTLSKQQMLMSVATKIAMQMNCKMGGDLWRVEIPLSNLMIIGID
                     SYHDSLTKGRSVLGFVASMNKSQTSFFSSCAFQHAQGEFGANLSTLMNNALKRYYQIN
                     EKFPERIIIFRDGVGDSQVNLVVDYELKQIKDTLDKVYPQGTVHKLAVVVVKKRINNR
                     FFANLRGGLSNPPPGTVIDDVVTKPHLYDYFIISQSVRQGSVSPTSYNVVYDTTGLKP
                     DHMQRLTYKLTHLYFNWPGTVRVPAPCMYAHKLAFLIGQSVHEVPNNRLADTLFYL"
BASE COUNT       822 a    806 c    796 g    633 t
ORIGIN
```

```
  Query Match             16.3%;  Score 663.8;  DB 3;  Length 3057;
  Best Local Similarity   57.5%;  Pred. No. 3.6e-111;
  Matches 1294;  Conservative    0;  Mismatches  940;  Indels   18;  Gaps    5;

Qy        533 CTGAGCACCAACCACTTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCAC 592
Db        505 CTGATAGCAAACGGCTTCAAGTTGAAGACTAAACCAGACTGGCAACTCTACCAGTACCGG 564

Qy        593 ATCGACTACAATCCCCTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCAT 652
Db        565 GTTGACTTTGAGCCTGAGATCCTGAACCCTAGAGCTCGTTTTGCCCTCCTGAAGGGCCAC 624

Qy        653 GAAGACCTCATTGGAAGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGA 712
Db        625 AGTGCTCTTCTTGGTAAAGGGTTAACGCTCGATATGGACACCATGTATTCTCTTTTCAAG 684

Qy        713 CTACAGCACAAGGTCACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATC 772
Db        685 CTAGCAGAAAAGGTAACCAACCTGAGCGCAGAGAGGAAGGATGGGTCTAACGTGAATATC 744

Qy        773 ACCATCACCCTGACCAACGAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAAC 832
Db        745 TGTGTGACTCATGTTGCTACTCTTAACCCCATGGCACCAAACACCCTGCATCTCTACAAC 804

Qy        833 ATCNTCTTCAGGAGGCTCTTGAAAATCATGAATTTGCAACAAATTGGACGGAATTATTAC 892
Db        805 GTCCTCTTTAGAAGGTGTTTGAAGATGATCAATATGGAGCAAGTTGGACGCAACTACTAT 864

Qy        893 AATCCAAGTGACCCGATTGATATTCCAAACCACAGGTTGGTGATCTGGCCCGGCTTCACC 952
Db        865 GATCCTACGGCCGCCATTGACATCAAGCAGCATGGGCTGCAGCTGTGGCCAGGGTTCGTG 924

Qy        953 ACCTCCATCCTTCAGTATGAGAACAACATCATGCTCTGCACAGACGTCAGCCACAAGGTG 1012
Db        925 ACATCCATCCTTCAGTACGAGTACGACGTGATGCTGCTGTCTGACATCAGCCACAAGGTT 984

Qy       1013 CTCCGCAGCGAGACTGTCCTAGACTTCATGTTCAATCTATACCAGCAGACAGAGGAGCAC 1072
Db        985 CTGCGTACCCAGACCGTCTGGGAGGTCATGAATGATCTCTTC---AACAAAGCTAGGGGC 1041

Qy       1073 AAGTTCCAGGAGCAAGTGTCGAAGGAGCTCATAGGCCTCATCGTTCTCACCAAGTACAAT 1132
Db       1042 AGGTTCAAGGAGGAAATCACAAAGCTCATGATTGGACAGATTGTCCTCACAAAGTACAAC 1101
```

```
Qy      1133 AACAAGACCTACCGGGTGGATGACATTGACTGGGACCAGAATCCAAAGAGCACCTTCAAG 1192
Db      1102 AACAAGACCTACCGGGTAGATGATATCGACTTTGACACTACTCCAGCGGATACTTTTGAG 1161

Qy      1193 AAGGCGGATGGCTCGGAGGTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAG 1252
Db      1162 A---CCCGTTCCGGACCAGTTTCCTATGTCGATTACTTCAAGAAGTCCTATGAGAGGGTC 1218

Qy      1253 ATCACGGACCTGAAGCAGCCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGC 1312
Db      1219 ATTCACGACGTGAACCAACCCATGCTGGTGTCTCGCCCAAAGAAGAGGGAGGAAAAGAAG 1278

Qy      1313 GGCACCCTGCCTGGCCCAGCTATGNTCATCCCTGAACTCTGCTATCTCACAGGCCTGACT 1372
Db      1279 GG------AGTTGGTCCAGCCTACCTACCGCCTGAGCTGTGCTTCCTGACTGGCCTAAGC 1332

Qy      1373 GATAAAATGCGCAATGATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACC 1432
Db      1333 GACGACATGCGCGCCGATTTCAACGTCATGAAGGACATGGGCCAACACACTCGTGTTGGA 1392

Qy      1433 CCTGAGCAGCGGCAGCGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAAT 1492
Db      1393 CCCCAAGACAGATGTAGAACGCTCAGCGGCTTCATCAAAAAGCTCAGCAGCAATGAAGAA 1452

Qy      1493 GTGCAGAGAGAGCTTCGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCT 1552
Db      1453 GTGAAGACTTACCTAGATTCCTGGGGCATGGAGTTCGACAAAGAGCAGGTCAAACTGACT 1512

Qy      1553 GGAAGAATCTTACAATCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCA 1612
Db      1513 GGGCGTGTCCTGCCACCTGAGAAACTATTCCAGAGGGGTAAACAGTTCAGTTACAATCCG 1572

Qy      1613 CAATTTGCAGACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTG 1672
Db      1573 AGCAATGCGGACTGGTCCAGAGACACCAGGGGTAATGCCCTCACTGATGCAAAGATCCTC 1632

Qy      1673 GATAACTGGCTGCTGATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAG 1732
Db      1633 AACAACTGGAAGATCTTCTACACACGCCGGGACGCCAACAGGGGCCAGGACTTCATCAAG 1692

Qy      1733 AACCTGTTCAAAGTGACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAG 1792
Db      1693 TCACTCGTACGAGTCGCAAACCCAATGGGTATGAACGTACGAGGACCAGAGATAGTGGAA 1752

Qy      1793 GT---GGATGACAGAACAGAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGAC 1849
Db      1753 CTTCCCGATGATCGTACAGAGACCTACACAAGATCCCTACAGGCACAAATCGCACAGGAT 1812

Qy      1850 ACTCAGATAGTTGTCTGTCTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAG 1909
Db      1813 ACACAGATTGTGGTGGTGATCCTGCCAACGAACAGAAAGGACCGGTACGATGCCATCAAG 1872

Qy      1910 AAGTACTTGTGTACAGACTGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGC 1969
Db      1873 AAGACGTGCGTGGTGACCCATCCATGCCCCAGCCAGGTCATCGTCTCCAGGACCCTGTCC 1932

Qy      1970 AAGCAGCAAACAGTCATGGCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATG 2029
Db      1933 AAGCAGCAGATGCTCATGTCTGTAGCCACCAAGATTGCCATGCAGATGAACTGCAAGATG 1992

Qy      2030 GGAGGCGAGCTCTGGCGGGTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGAC 2089
Db      1993 GGCGGAGACCTCTGGCGAGTCGAGATCCCGCTGAGCAACCTGATGATCATCGGTATTGAC 2052

Qy      2090 TGTTACCATGACACCACAGCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAAT 2149
Db      2053 AGCTACCATGACTCTCTGACCAAAGGCCGCTCCGTCCTGGGATTCGTGGCTTCCATGAAC 2112

Qy      2150 GAAGGGATGACCCGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTG 2209
Db      2113 AAGTCTCAAACCAGCTTCTTCTCTAGTTGTGCATTCCAGCACGCACAGGGAGAATTCGGT 2172

Qy      2210 GATGGTCTCAAGGTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATAC 2269
Db      2173 GCTAACCTGAGTACCCTCATGAACAATGCCCTGAAGCGATACTATCAGATCAACGAGAAG 2232

Qy      2270 ATGCCCAGCCGTGTCATCGTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTG 2329
Db      2233 TTTCCTGAGCGTATCATCATCTTCCGTGATGGTGTTGGTGACAGCCAAGTCAATCTGGTC 2292

Qy      2330 GTCAATTATGAGGTCCCACAGTTCCTAGATTGCCTCAAGTCAGTCGGGAG---AGGTTAC 2386
Db      2293 GTTGACTATGAGCTCAAACAAATCAAGGACACCTTGGACAAAGTCTACCCTCAAGGGACG 2352

Qy      2387 AACCCAAGACTGACTGTAATCGTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAG 2446
Db      2353 GTTCACAAGCTGGCCGTTGTGGTCGTCAAGAAGCGCATCAATAACCGCTTCTTTGCCAAT 2412

Qy      2447 TCTGGGGGAAGACTTCAGAACCCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGA 2506
Db      2413 CTGAGAGGAGGCCTGTCCAACCCTCCCCCTGGAACCGTCATCGATGATGTCGTCACCAAG 2472

Qy      2507 CCAGAGTGGTATGACTTTTTCATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCA 2566
Db      2473 CCCCACCTCTATGACTACTTCATCATCAGTCAGAGTGTCCGCCAGGGCTCCGTGTCCCCC 2532

Qy      2567 ACACACTACAATGTCATCTATGACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTG 2626
Db      2533 ACGTCCTACAACGTTGTCTATGACACCACTGGTCTCAAGCCAGACCACATGCAGAGGTTG 2592

Qy      2627 ACATACAAGNTCTGCCACGTGTACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCT 2686
Db      2593 ACCTACAAGCTGACCCATCTCTACTTCAACTGGCCGGGAACGGTGCGTGTTCCTGCCCCG 2652

Qy      2687 TGCCAGTATGCACACAAGCTGGCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAAC 2746
Db      2653 TGCATGTATGCACACAAGCTGGCATTCCTGATCGGGCAGAGCGTCCATGAAGTGCCCAAC 2712

Qy      2747 CTCTCCCTGTCCAACCGCCTCTACTACCTCTA 2778
Db      2713 AACAGACTAGCCGACACCCTCTTCTATCTCTA 2744
```

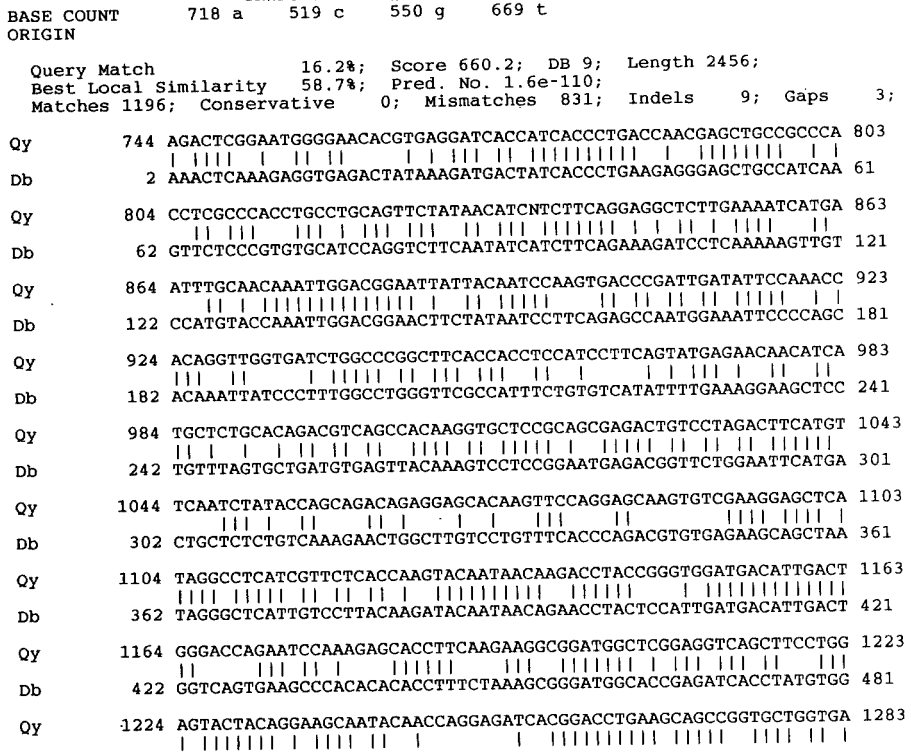
```
RESULT 13
AK096837
LOCUS       AK096837               2456 bp    mRNA    linear   PRI 15-JUL-2002
DEFINITION  Homo sapiens cDNA FLJ39518 fis, clone PUAEN2001188, moderately
            similar to Mus musculus mRNA for MIWI (piwi).
ACCESSION   AK096837
VERSION     AK096837.1  GI:21756416
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens pulmonary artery endothelial cells (HPAEC) cDNA to
            mRNA, clone_lib:PUAEN2 clone:PUAEN2001188.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Tanigami,A., Fujiwara,T., Shibahara,T., Goto,Y., Hirao,M.,
            Shimizu,F., Wakebe,H., Ono,T., Hishigaki,H., Watanabe,T., Ozaki,K.,
            Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
            Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
            Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
            Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
            Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
            Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2456)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
```

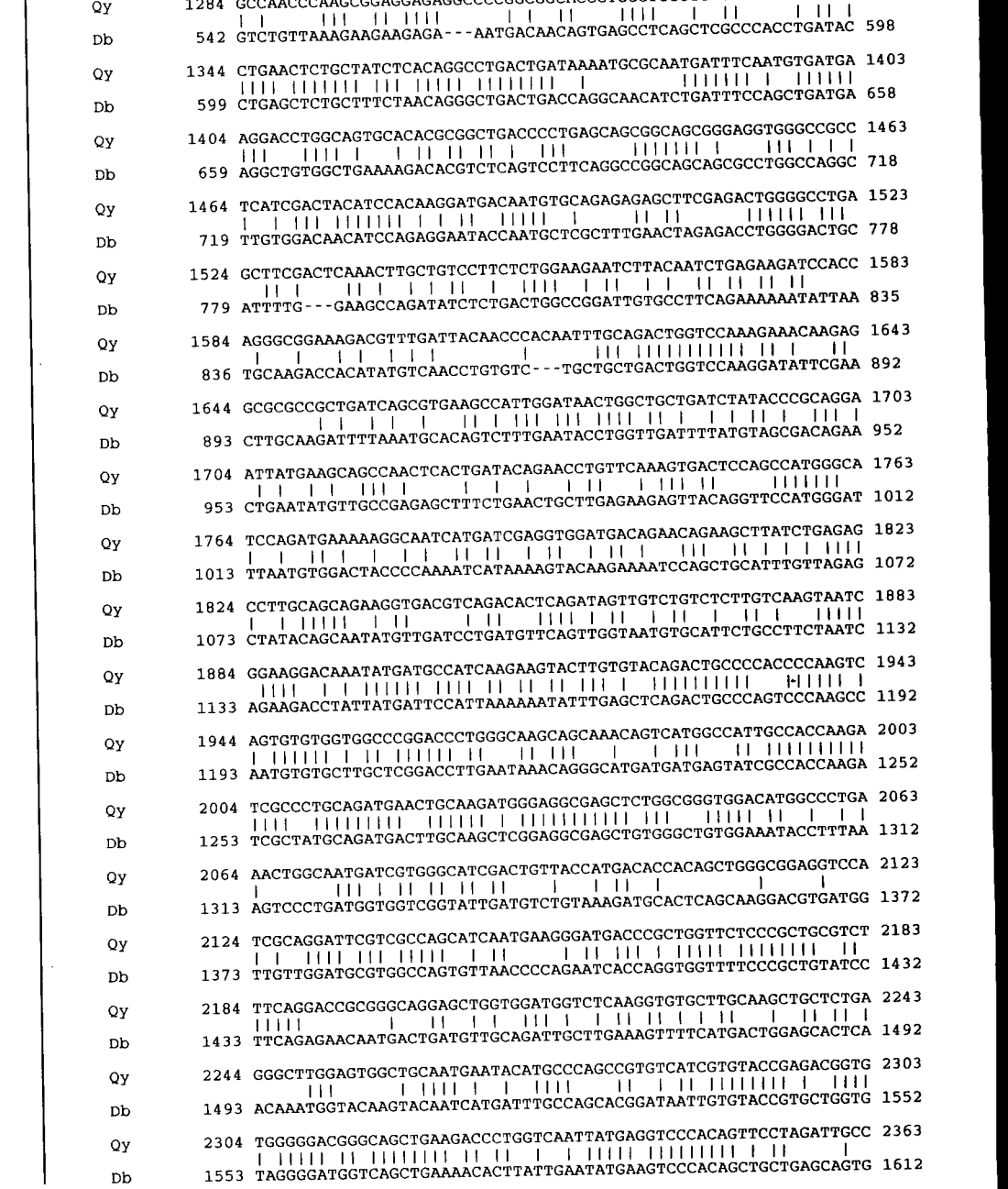
Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES Location/Qualifiers
source 1. .2456
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PUAEN2001188"
/cell_type="pulmonary artery endothelial cells (HPAEC)"
/clone_lib="PUAEN2"
/note="cloning vector: pME18SFL3-primary culture, pulmonary artery endothelial cells"
CDS 28. .2028
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC04873.1"
/db_xref="GI:21756417"
/translation="MTITLKRELPSSSPVCIQVFNIIFRKILKKLSMYQIGRNFYNPS
EPMEIPQHKLSLWPGFAISVSYFERKLLFSADVSYKVLRNETVLEFMTALCQRTGLSC
FTQTCEKQLIGLIVLTRYNNRTYSIDDIDWSVKPTHTFLKRDGTEITYVDYYKQQYDI
TVSDLNQPMLVSLLKKKRNDNSEPQLAHLIPELCFLTGLTDQATSDFQLMKAVAEKTR
LSPSGRQQRLARLVDNIQRNTNARFELETWGLHFGSQISLTGRIVPSEKILMQDHICQ
PVSAADWSKDIRTCKILNAQSLNTWLILCSDRTEYVAESFLNCLRRVTGSMGFNVDYP
KIIKVQENPAAFVRAIQQYVDPDVQLVMCILPSNQKTYYDSIKKYLSSDCPVPSQCVL
ARTLNKQGMMMSIATKIAMQMTCKLGGELWAVEIPLKSLMVVGIDVCKDALSKDVMVV
GCVASVNPRITRWFSRCILQRTMTDVADCLKVFMTGALNKWYKYNHDLPARIIVYRAG
VGDGQLKTLIEYEVPQLLSSVAESSSNTSSRLSVIVVRKKCMPRFFTEMNRTVQNPPL
GTVVDSEATRNEWYDFYLISQVACRGTVSPTYYNVIYDDNGLKPDHMQRLTFKLCHLY
YNWPGIVSVPAPCQYAHKLTFLVAQSIHKEPSLELANHLFYL"

BASE COUNT 718 a 519 c 550 g 669 t
ORIGIN

Query Match 16.2%; Score 660.2; DB 9; Length 2456;
Best Local Similarity 58.7%; Pred. No. 1.6e-110;
Matches 1196; Conservative 0; Mismatches 831; Indels 9; Gaps 3;





```
Qy       2364 TCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATCGTGGTGAAGAAGCGTG 2423
Db       1613 TGGCAGAATCCAGCTCAAATACCAGCTCAAGACTGTCGGTGATTGTGGTCAGGAAGAAGT 1672

Qy       2424 TCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAACCCTCTTCCAGGGACAG 2483
Db       1673 GCATGCCACGATTCTTTACCGAAATGAACCGCACTGTACAGAACCCCCCACTTGGCACTG 1732

Qy       2484 TCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTCATCGTGAGCCAGGCAG 2543
Db       1733 TTGTGGATTCAGAAGCAACACGTAACGAATGGTATGACTTTTATCTGATCAGCCAGGTGG 1792

Qy       2544 TGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTATGACAGCAGTGGCCTGA 2603
Db       1793 CCTGCCGGGGAACTGTTAGTCCTACCTACTATAATGTCATCTATGATGACAACGGCTTGA 1852

Qy       2604 AGCCCGACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTGTACTATAATTGGCCTG 2663
Db       1853 AGCCCGACCATATGCAGAGACTTACATTCAAATTGTGCCACCTGTACTACAACTGGCCGG 1912

Qy       2664 GAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTGGCCTTCCTCGTGGGCC 2723
Db       1913 GCATAGTCAGTGTCCCAGCACCATGTCAGTATGCTCACAAGCTGACCTTTCTGGTGGCAC 1972

Qy       2724 AGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTCTACTACCTCTAA 2779
Db       1973 AAAGCATTCATAAAGAACCCAGTCTGGAATTAGCCAACCATCTCTTCTACCTGTGA 2028
```

RESULT 14
AY014900

```
LOCUS       AY014900                2605 bp    mRNA    linear   INV 02-JAN-2001
DEFINITION  Strongylocentrotus purpuratus clone 6-2 seawi mRNA, partial cds.
ACCESSION   AY014900
VERSION     AY014900.1  GI:12007641
KEYWORDS    .
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 2605)
  AUTHORS   Rodriguez,A.J. and Bonder,E.M.
  TITLE     Seawi - Cloning and Molecular Characterization of a Sea Urchin
            Homolog of Piwi
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2605)
  AUTHORS   Rodriguez,A.J. and Bonder,E.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-2000) Biological Sciences, Rutgers
            University-Newark, 101 Warren Street Room 135, Newark, NJ 07102,
            USA
FEATURES             Location/Qualifiers
     source          1. .2605
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="6-2"
     CDS             181. .>2605
                     /codon_start=1
                     /product="seawi"
                     /protein_id="AAG42534.1"
                     /db_xref="GI:12007642"
                     /translation="MDRRPGEQAPPPMQQAAPPQEASSGMPVPGRGRSRGGTTQPGPG
                     AVPKQEPRGELTSQIAKLAIGASGEGRVGGAVQDRRRRRDEIELEPTTRPENFVKQAI
                     AGDKIALIANGFKLKTKPDWQLYQYRVDFEPEILNPRARFALLKGHSALLGKGLTLDM
                     DTMYSLFKLAEKVTNLSAERKDGSNVNICVTHVATLNPMAPNTLHLYNVLFRRCLKMI
                     NMEQVGRNYYDPTAAIDIKQHGLQLWPGFVTSILQYEYDVMLLSDISHKVLRTQTVWE
                     VMNDLFNKARGRFKEEITKLMIGQIVLTKYNNKTYRVDDIDFDTTPADTFETRSGPVS
                     YVDYFKKSYERVIHDVNQPMLVSRPKKREEKKGVGPAYLPPELCFLTGLSDDMRADFN
                     VMKDMGQHTRVGPQDRCRTLSGFIKKLSSNEEVKTYLDSWGMEFDKEQVKLTGRVLPP
                     EKLFQRGKQFSYNPSNADWSRDTRGNALTDAKILNNWKIFYTRRDANRGQDFIKSLVR
                     VANPMGMNVRGPEIVELPDDRTETYTRSLQAQIAQDTQIVVVILPTNRKDRYDAIKKT
                     CVVTHPCPSQVIVSRTLSKQQMLMSVATKIAMQMNCKMGGDLWRVEIPLSNLMIIGID
                     SYHDSLTKGRSVLGFVASMNKSQTSFFSSCAFQHAQGEFGANLSTLMNNALKRYYQIN
                     EKFPERIIIFRDGVGDSQVNLVVDYELKQIKDTLDKVYPQGTVHKLAVVVVKKRINNR
                     FFANLRGGLSNPPPGTVIDDVVTKPHLYDYFIISQSVRQGSVSPTSYNVVYDTTGLKP
                     DHMQRLTYGI"
BASE COUNT      718 a    695 c    686 g    506 t
ORIGIN
```

Query Match 14.5%; Score 589.2; DB 3; Length 2605;
Best Local Similarity 56.6%; Pred. No. 1.4e-97;
Matches 1193; Conservative 0; Mismatches 895; Indels 18; Gaps 5;

```
Qy        533 CTGAGCACCAACCACTTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCAC 592
Db        505 CTGATAGCAAACGGCTTCAAGTTGAAGACTAAACCAGACTGGCAACTCTACCAGTACCGG 564

Qy        593 ATCGACTACAATCCCCTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCAT 652
Db        565 GTTGACTTTGAGCCTGAGATCCTGAACCCTAGAGCTCGTTTTGCCCTCCTGAAGGGCCAC 624

Qy        653 GAAGACCTCATTGGAAGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGA 712
Db        625 AGTGCTCTTCTTGGTAAAGGGTTAACGCTCGATATGGACACCATGTATTCTCTTTTCAAG 684

Qy        713 CTACAGCACAAGGTCACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATC 772
Db        685 CTAGCAGAAAAGGTAACCAACCTGAGCGCAGAGAGGAAGGATGGGTCTAACGTGAATATC 744

Qy        773 ACCATCACCCTGACCAACGAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAAC 832
Db        745 TGTGTGACTCATGTTGCTACTCTTAACCCCATGGCACCAAACACCCTGCATCTCTACAAC 804

Qy        833 ATCNTCTTCAGGAGGCTCTTGAAAATCATGAATTTGCAACAAATTGGACGGAATTATTAC 892
Db        805 GTCCTCTTTAGAAGGTGTTTGAAGATGATCAATATGGAGCAAGTTGGACGCAACTACTAT 864

Qy        893 AATCCAAGTGACCCGATTGATATTCCAAACCACAGGTTGGTGATCTGGCCCGGCTTCACC 952
Db        865 GATCCTACGGCCGCCATTGACATCAAGCAGCATGGGCTGCAGCTGTGGCCAGGGTTCGTG 924

Qy        953 ACCTCCATCCTTCAGTATGAGAACAACATCATGCTCTGCACAGACGTCAGCCACAAGGTG 1012
Db        925 ACATCCATCCTTCAGTACGAGTACGACGTGATGCTGCTGTCTGACATCAGCCACAAGGTT 984

Qy       1013 CTCCGCAGCGAGACTGTCCTAGACTTCATGTTCAATCTATACCAGCAGACAGAGGAGCAC 1072
Db        985 CTGCGTACCCAGACCGTCTGGGAGGTCATGAATGATCTCTTC---AACAAAGCTAGGGGC 1041

Qy       1073 AAGTTCCAGGAGCAAGTGTCGAAGGAGCTCATAGGCCTCATCGTTCTCACCAAGTACAAT 1132
Db       1042 AGGTTCAAGGAGGAAATCACAAAGCTCATGATTGGACAGATTGTCCTCACAAAGTACAAC 1101

Qy       1133 AACAAGACCTACCGGGTGGATGACATTGACTGGGACCAGAATCCAAAGAGCACCTTCAAG 1192
Db       1102 AACAAGACCTACCGGGTAGATGATATCGACTTTGACACTACTCCAGCGGATACTTTTGAG 1161

Qy       1193 AAGGCGGATGGCTCGGAGGTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAG 1252
Db       1162 A---CCCGTTCCGGACCAGTTTCCTATGTCGATTACTTCAAGAAGTCCTATGAGAGGGTC 1218

Qy       1253 ATCACGGACCTGAAGCAGCCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGC 1312
Db       1219 ATTCACGACGTGAACCAACCCATGCTGGTGTCTCGCCCAAAGAAGAGGGAGGAAAAGAAG 1278

Qy       1313 GGCACCCTGCCTGGCCCAGCTATGNTCATCCCTGAACTCTGCTATCTCACAGGCCTGACT 1372
Db       1279 GG------AGTTGGTCCAGCCTACCTACCGCCTGAGCTGTGCTTCCTGACTGGCCTAAGC 1332

Qy       1373 GATAAAATGCGCAATGATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACC 1432
Db       1333 GACGACATGCGCGCCGATTTCAACGTCATGAAGGACATGGGCCAACACACTCGTGTTGGA 1392

Qy       1433 CCTGAGCAGCGGCAGCGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAAT 1492
```

```
Db       1393 CCCCAAGACAGATGTAGAACGCTCAGCGGCTTCATCAAAAAGCTCAGCAGCAATGAAGAA 1452

Qy       1493 GTGCAGAGAGAGCTTCGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCT 1552

Db       1453 GTGAAGACTTACCTAGATTCCTGGGGCATGGAGTTCGACAAAGAGCAGGTCAAACTGACT 1512

Qy       1553 GGAAGAATCTTACAATCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCA 1612

Db       1513 GGGCGTGTCCTGCCACCTGAGAAACTATTCCAGAGGGGTAAACAGTTCAGTTACAATCCG 1572

Qy       1613 CAATTTGCAGACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTG 1672

Db       1573 AGCAATGCGGACTGGTCCAGAGACACCAGGGGTAATGCCCTCACTGATGCAAAGATCCTC 1632

Qy       1673 GATAACTGGCTGCTGATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAG 1732

Db       1633 AACAACTGGAAGATCTTCTACACACGCCGGGACGCCAACAGGGGCCAGGACTTCATCAAG 1692

Qy       1733 AACCTGTTCAAAGTGACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAG 1792

Db       1693 TCACTCGTACGAGTCGCAAACCCAATGGGTATGAACGTACGAGGACCAGAGATAGTGGAA 1752

Qy       1793 GTG---GATGACAGAACAGAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGAC 1849

Db       1753 CTTCCCGATGATCGTACAGAGACCTACACAAGATCCCTACAGGCACAAATCGCACAGGAT 1812

Qy       1850 ACTCAGATAGTTGTCTGTCTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAG 1909

Db       1813 ACACAGATTGTGGTGGTGATCCTGCCAACGAACAGAAAGGACCGGTACGATGCCATCAAG 1872

Qy       1910 AAGTACTTGTGTACAGACTGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGC 1969

Db       1873 AAGACGTGCGTGGTGACCCATCCATGCCCCAGCCAGGTCATCGTCTCCAGGACCCTGTCC 1932

Qy       1970 AAGCAGCAAACAGTCATGGCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATG 2029

Db       1933 AAGCAGCAGATGCTCATGTCTGTAGCCACCAAGATTGCCATGCAGATGAACTGCAAGATG 1992

Qy       2030 GGAGGCGAGCTCTGGCGGGTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGAC 2089

Db       1993 GGCGGAGACCTCTGGCGAGTCGAGATCCCGCTGAGCAACCTGATGATCATCGGTATTGAC 2052

Qy       2090 TGTTACCATGACACCACAGCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAAT 2149

Db       2053 AGCTACCATGACTCTCTGACCAAAGGCCGCTCCGTCCTGGGATTCGTGGCTTCCATGAAC 2112

Qy       2150 GAAGGGATGACCCGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTG 2209

Db       2113 AAGTCTCAAACCAGCTTCTTCTCTAGTTGTGCATTCCAGCACGCACAGGGAGAATTCGGT 2172

Qy       2210 GATGGTCTCAAGGTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATAC 2269

Db       2173 GCTAACCTGAGTACCCTCATGAACAATGCCCTGAAGCGATACTATCAGATCAACGAGAAG 2232

Qy       2270 ATGCCCAGCCGTGTCATCGTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTG 2329

Db       2233 TTTCCTGAGCGTATCATCATCTTCCGTGATGGTGTTGGTGACAGCCAAGTCAATCTGGTC 2292

Qy       2330 GTCAATTATGAGGTCCCACAGTTCCTAGATTGCCTCAAGTCAGT---CGGGAGAGGTTAC 2386

Db       2293 GTTGACTATGAGCTCAAACAAATCAAGGACACCTTGGACAAAGTCTACCCTCAAGGGACG 2352

Qy       2387 AACCCAAGACTGACTGTAATCGTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAG 2446

Db       2353 GTTCACAAGCTGGCCGTTGTGGTCGTCAAGAAGCGCATCAATAACCGCTTCTTTGCCAAT 2412

Qy       2447 TCTGGGGGAAGACTTCAGAACCCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGA 2506

Db       2413 CTGAGAGGAGGCCTGTCCAACCCTCCCCCTGGAACCGTCATCGATGATGTCGTCACCAAG 2472

Qy       2507 CCAGAGTGGTATGACTTTTTCATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCA 2566

Db       2473 CCCCACCTCTATGACTACTTCATCATCAGTCAGAGTGTCCGCCAGGGCTCCGTGTCCCCC 2532

Qy       2567 ACACACTACAATGTCATCTATGACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTG 2626

Db       2533 ACGTCCTACAACGTTGTCTATGACACCACTGGTCTCAAGCCAGACCACATGCAGAGGTTG 2592

Qy       2627 ACATAC 2632

Db       2593 ACCTAC 2598
```

RESULT 15
AC116715
LOCUS AC116715 60131 bp DNA linear HTG 05-JUL-2002
DEFINITION Mus musculus clone RP23-129N14, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC116715
VERSION AC116715.3 GI:21699658
KEYWORDS HTG; HTGS_PHASE0.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 60131)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP23-129N14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 60131)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 60131)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Jul 5, 2002 this sequence version replaced gi:20800207.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L22417
Center clone name: 129_N_14
----------.
* NOTE: This record contains 75 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 674: contig of 674 bp in length
* 675 774: gap of 100 bp
* 775 1459: contig of 685 bp in length
* 1460 1559: gap of 100 bp
* 1560 2262: contig of 703 bp in length
* 2263 2362: gap of 100 bp
* 2363 3068: contig of 706 bp in length
* 3069 3168: gap of 100 bp
* 3169 3881: contig of 713 bp in length
* 3882 3981: gap of 100 bp
* 3982 4706: contig of 725 bp in length
* 4707 4806: gap of 100 bp
* 4807 5509: contig of 703 bp in length
* 5510 5609: gap of 100 bp
* 5610 6292: contig of 683 bp in length
* 6293 6392: gap of 100 bp
* 6393 7083: contig of 691 bp in length
* 7084 7183: gap of 100 bp
* 7184 7892: contig of 709 bp in length
* 7893 7992: gap of 100 bp
* 7993 8700: contig of 708 bp in length
* 8701 8800: gap of 100 bp
* 8801 9480: contig of 680 bp in length
* 9481 9580: gap of 100 bp
* 9581 10302: contig of 722 bp in length
* 10303 10402: gap of 100 bp
* 10403 11101: contig of 699 bp in length
* 11102 11201: gap of 100 bp
* 11202 11895: contig of 694 bp in length
* 11896 11995: gap of 100 bp
* 11996 12681: contig of 686 bp in length
* 12682 12781: gap of 100 bp
* 12782 13474: contig of 693 bp in length
* 13475 13574: gap of 100 bp
* 13575 14277: contig of 703 bp in length
* 14278 14377: gap of 100 bp
* 14378 15076: contig of 699 bp in length
* 15077 15176: gap of 100 bp
* 15177 15901: contig of 725 bp in length
* 15902 16001: gap of 100 bp
* 16002 16715: contig of 714 bp in length
* 16716 16815: gap of 100 bp
* 16816 17548: contig of 733 bp in length
* 17549 17648: gap of 100 bp
* 17649 18350: contig of 702 bp in length
* 18351 18450: gap of 100 bp
* 18451 19148: contig of 698 bp in length
* 19149 19248: gap of 100 bp
* 19249 19956: contig of 708 bp in length
* 19957 20056: gap of 100 bp
* 20057 20754: contig of 698 bp in length
* 20755 20854: gap of 100 bp
* 20855 21539: contig of 685 bp in length
* 21540 21639: gap of 100 bp
* 21640 22336: contig of 697 bp in length
* 22337 22436: gap of 100 bp
* 22437 23135: contig of 699 bp in length
* 23136 23235: gap of 100 bp
* 23236 23958: contig of 723 bp in length
* 23959 24058: gap of 100 bp
* 24059 24737: contig of 679 bp in length
* 24738 24837: gap of 100 bp
* 24838 25557: contig of 720 bp in length
* 25558 25657: gap of 100 bp
* 25658 26372: contig of 715 bp in length
* 26373 26472: gap of 100 bp
* 26473 27142: contig of 670 bp in length
* 27143 27242: gap of 100 bp
* 27243 27953: contig of 711 bp in length
* 27954 28053: gap of 100 bp
* 28054 28744: contig of 691 bp in length
* 28745 28844: gap of 100 bp
* 28845 29557: contig of 713 bp in length
* 29558 29657: gap of 100 bp
* 29658 30363: contig of 706 bp in length
* 30364 30463: gap of 100 bp
* 30464 31162: contig of 699 bp in length
* 31163 31262: gap of 100 bp
* 31263 31973: contig of 711 bp in length
* 31974 32073: gap of 100 bp
* 32074 32792: contig of 719 bp in length
* 32793 32892: gap of 100 bp
* 32893 33592: contig of 700 bp in length
* 33593 33692: gap of 100 bp
* 33693 34411: contig of 719 bp in length
* 34412 34511: gap of 100 bp
* 34512 35233: contig of 722 bp in length
* 35234 35333: gap of 100 bp
* 35334 36029: contig of 696 bp in length
* 36030 36129: gap of 100 bp
* 36130 36837: contig of 708 bp in length
* 36838 36937: gap of 100 bp
* 36938 37636: contig of 699 bp in length
* 37637 37736: gap of 100 bp
* 37737 38426: contig of 690 bp in length
* 38427 38526: gap of 100 bp
* 38527 39241: contig of 715 bp in length
* 39242 39341: gap of 100 bp
* 39342 40043: contig of 702 bp in length
* 40044 40143: gap of 100 bp
* 40144 40859: contig of 716 bp in length
* 40860 40959: gap of 100 bp
* 40960 41658: contig of 699 bp in length
* 41659 41758: gap of 100 bp
* 41759 42443: contig of 685 bp in length
* 42444 42543: gap of 100 bp
* 42544 43242: contig of 699 bp in length
* 43243 43342: gap of 100 bp
* 43343 44041: contig of 699 bp in length
* 44042 44141: gap of 100 bp
* 44142 44858: contig of 717 bp in length
* 44859 44958: gap of 100 bp
* 44959 45647: contig of 689 bp in length
* 45648 45747: gap of 100 bp
* 45748 46455: contig of 708 bp in length

Query Match 14.3%; Score 579.8; DB 2; Length 60131;
Best Local Similarity 81.3%; Pred. No. 8.4e-96;
Matches 612; Conservative 1; Mismatches 135; Indels 5; Gaps 1;

Qy 3110 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGCATGTTTGAAT 3169

```
Db      52265 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGCATGTTTGAAT 52324

Qy       3170 TTAAAAGAGGAAAGGACACAAGGCAAGAATGCCCAGAGCATCCTAGATGGGTGCTATATG 3229

Db      52325 TTAAAAGAGGAAAGGACACAAGGCAAGAATGCCCAGAGCATCCTAGATGGGTGCTATATG 52384

Qy       3230 GGACTCTGACACCTCAGGTTCTTAAAGATAAAAAGGCCAGAGGTACCTTTCTGTAATTCT 3289

Db      52385 GGACTCTGACACCTCAGGTTCTTAAAGATAAAAAGGCCAGAGGTACCTTTCTGTAATTCT 52444

Qy       3290 TTTAAGAGTTAGGTCATAAGGAATGAAAAGAGTTACTTTTCTCCCTCCATGAAAGTGTTT 3349

Db      52445 TTTAAGAGTTAGGTCATAAGGAATGAAAAGAGTTACTTTTCTCCCTCCATGAAAGTGTTT 52504

Qy       3350 TTTAAGACTGACAAGGCCGGTGGTGGAGTGAGTGTTGTTGTTTTGTTTCGCAATTTTCTT 3409

Db      52505 TTTAAGACTGACAAGGCCGGTGGTGGAGTGAGTGTTGTTGTTTTGTTTCGCAATTTTCTT 52564

Qy       3410 ACCAAGTGAAATTACTGACCATAATTTTATTTAATAGTTTTTAGTTTTTGAAAGCCTGTG 3469

Db      52565 ACCAAGTGAAATTACTGACCATAATTTTATTTAATAGGTGTTAGTTTTTGAAAGCCTGTG 52624

Qy       3470 TTTTCTTTGCTTTTAAACTACAGGCGACCACTCTGGGATAGGATTCCACTCCCAAAAGTA 3529

Db      52625 TTTTCTTTGCTTTTAAACTACAGGCGACCACTCTGGGATAGGATTCCACTCCCAAAAGTA 52684

Qy       3530 ATGTGTTGGGTGTGGTGGCACACGCCTGTGATCTCAGGACTCAGGAGGTGAAGCCAGGGG 3589

Db      52685 ATGTGTTGGGTGTGGTGGCACACGCCTGTGATCTCAGGACTCAGGAGGTGAAGCCAGGGG 52744

Qy       3590 AACCCCAAGTTCCCGGTCAGCCTAGGCTACATAGTGAGACTTTGTCTTAACAAACAGAAA 3649

Db      52745 AACCCCAAGTTCCCGGACAGCCTATGCTACATAGTGAGACTTTGTCTTACCA-----ACA 52799

Qy       3650 AAAAGAAAAAGAAAGAAAAAAAAAAAGAAAGAAAAAAAAAARAAAGAAAAATGGAGCAGTAT 3709

Db      52800 GAAAAAAGAAACGATAGAAAAAACAAGAAAGAAGCCCNAGATGGTATTTTGGAGCGGTAT 52859

Qy       3710 GAGTTATGGGTGCCAGTTTCTGCAGGAGACTTTGGGAGTAAGAGCTGCCTCTGTTCCTAA 3769

Db      52860 GAGTTATGGGTGCCAGTTCCTGCACGAGACTTTGGGAGAAAGAGCTGCCTCNNNNNNNNN 52919

Qy       3770 GAGGCTGCAAGGCTGCGGCTGGGGGGCCAGTTCAGGTGCCAGGTGAGGTACGGCCACATC 3829

Db      52920 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 52979

Qy       3830 TATTTGCCTGGCTCTTTGACAAAGAGAGCAGCT 3862

Db      52980 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNGT 53012
```

Search completed: July 22, 2003, 06:18:52
Job time : 10264.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2003, 11:38:12 ; Search time 810.65 Seconds
(without alignments)
11289.862 Million cell updates/sec

Title: US-09-873-737A-3
Perfect score: 4064
Sequence: 1 agaagcctgctacaggggt.....aaaaaaaaaaaaaaaaaaa 4064

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4059.6	99.9	4064	21	AAA07587	Mouse piwi gene, d
	2	1833.6	45.1	3472	21	AAA07588	Human piwi gene, d
	3	399.4	9.8	3649	22	AAS13630	cDNA sequence enco
	4	345	8.5	2272	22	AAH15960	Human cDNA sequenc
	5	318.2	7.8	3524	23	ABL10535	Drosophila melanog
	6	311.8	7.7	423	21	AAF21880	Human breast and o
	7	311.4	7.7	3047	21	AAA07586	Drosophila piwi ge
	8	286	7.0	2838	23	ABL10571	Drosophila melanog
	9	259.6	6.4	1733	22	AAH14032	Human cDNA sequenc
	10	231.2	5.7	300	20	AAZ13404	Human gene express
	11	185.4	4.6	1026	23	AAS70144	DNA encoding novel
	12	175	4.3	5671	23	ABL10570	Drosophila melanog
	13	159.4	3.9	7667	23	ABL10534	Drosophila melanog
	14	157.6	3.9	548	24	ABN62610	Human cancer relat
	15	127.4	3.1	389	22	AAI88514	Human polynucleoti
	16	124	3.1	1954	21	AAA23441	cDNA encoding huma
	17	123.8	3.0	390	22	AAI85202	Human polynucleoti
	18	123.8	3.0	1225	23	ABK72084	Human cDNA encodin
	19	123.8	3.0	1225	24	ABK91676	cDNA encoding nove
	20	123.8	3.0	2288	22	AAH33398	Human colon cancer
	21	123.8	3.0	2288	22	AAF72807	Secreted protein g
	22	123	3.0	300	20	AAZ14372	Human gene express
	23	122.4	3.0	2262	21	AAZ52560	Human secreted pro
	24	122.2	3.0	1376	20	AAX04325	Human secreted pro
c	25	122	3.0	396	22	AAF94825	Human ovarian canc
c	26	122	3.0	396	24	ABT03092	Human ovarian carc
c	27	122	3.0	396	24	ABL48775	Ovarian carcinoma
c	28	121.8	3.0	5845	24	ABL33662	Human immune syste
c	29	121.6	3.0	425	23	ABV38062	Human prostate exp
	30	121.6	3.0	463	22	AAI83750	Human polynucleoti
	31	121.6	3.0	671	22	AAD05318	Human secreted pro
c	32	121.6	3.0	8946	24	ABL32911	Human immune syste
	33	121.4	3.0	439	23	ABV57512	Human prostate exp
	34	121.4	3.0	497	23	ABV58273	Human prostate exp
	35	121.4	3.0	1933	21	AAA23442	cDNA encoding huma
c	36	121.2	3.0	6668	24	ABL33696	Human immune syste
	37	120.6	3.0	580	23	ABV58017	Human prostate exp
	38	120.6	3.0	887	21	AAC59297	Human secreted pro
c	39	120.6	3.0	15832	22	AAS45399	Chemically pretrea
c	40	120.6	3.0	15832	24	ABL33343	Human immune syste
c	41	120.6	3.0	15832	24	ABK28244	DNA transcription
c	42	120.4	3.0	375	23	ABV44911	Human prostate exp
	43	120.4	3.0	403	22	AAI83197	Human polynucleoti
	44	120.4	3.0	528	23	ABV58464	Human prostate exp
	45	120.4	3.0	1798	24	AAS20587	Human methionine a

ALIGNMENTS

RESULT 1
AAA07587
ID AAA07587 standard; DNA; 4064 BP.
XX
AC AAA07587;
XX
DT 29-AUG-2000 (first entry)
XX
DE Mouse piwi gene, designated miwi.
XX
KW Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW anaemia; immunodeficiency; male infertility; mouse; ds.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 191..2779
FT /*tag= a
FT /product= miwi
FT /transl_except= (pos:450..452; aa:Xaa)
FT /transl_except= (pos:1337..1339; aa:Xaa)
FT /transl_except= (pos:2636..2638; aa:Xaa)
FT /note= "Xaa= Leu or Ile"
FT /transl_except= (pos:836..838; aa:Xaa)
FT /note= "Xaa= unspecified amino acid"
XX
PN WO200032039-A1.
XX
PD 08-JUN-2000.
XX
PF 03-DEC-1999; 99WO-US28764.
XX

PR 04-DEC-1998; 98US-0110901.
XX
PA (UYDU-) UNIV DUKE.
XX
PI Lin H;
XX
DR WPI; 2000-412085/35.
DR P-PSDB; AAY90234.
XX
PT Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT therapy of diseases such as cancer and in various research and
PT diagnostic applications -
XX
PS Claim 19; Page 180-185; 201pp; English.
XX
CC This sequence encodes the mouse piwi family protein, designated
CC miwi. The piwi family nucleic acids and polypeptides are used in gene
CC therapy of diseases such as cancer and also in various research and
CC diagnostic applications. The sequences can also be used to treat
CC tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ Sequence 4064 BP; 1114 A; 978 C; 1077 G; 890 T; 5 other;

Query Match 99.9%; Score 4059.6; DB 21; Length 4064;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4064; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGAAGCCTGCTACAGGGGGTGGCGTGAGGAGCCTCAAGGACTGCGGACCTAAGCGTGCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGAAGCCTGCTACAGGGGGTGGCGTGAGGAGCCTCAAGGACTGCGGACCTAAGCGTGCTA 60

Qy         61 AGCACGGGGCCAGGATAGGACTGGACGCTGCTGGACCACGACGATCAGGGAGTGACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGCACGGGGCCAGGATAGGACTGGACGCTGCTGGACCACGACGATCAGGGAGTGACCCCC 120

Qy        121 ACCAGCATCCAGCCTGCGGCAGTTGTAGGACTGGAAGGAGGCCAAGCTGAACACCTGAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCAGCATCCAGCCTGCGGCAGTTGTAGGACTGGAAGGAGGCCAAGCTGAACACCTGAGC 180

Qy        181 TGACTGGAAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGACTGGAAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGA 240

Qy        241 GACGGTGCAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GACGGTGCAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACC 300

Qy        301 TCAACAGTCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCAACAGTCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGT 360

Qy        361 CGGAGCCACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGAGCCACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACT 420

Qy        421 GGCAGAGAGAGGAGGGCGTCGCCGAGACTTCCATGACNTTGGTGTGAACACCAGACAGAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCAGAGAGAGGAGGGCGTCGCCGAGACTTCCATGACNTTGGTGTGAACACCAGACAGAA 480

Qy        481 CCTTGACCATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCTTGACCATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCAC 540

Qy        541 CAACCACTTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAACCACTTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTA 600

Qy        601 CAATCCCCTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAATCCCCTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCT 660

Qy        661 CATTGGAAGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CATTGGAAGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCA 720

Qy        721 CAAGGTCACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCAC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CAAGGTCACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCAC 780

Qy        781 CCTGACCAACGAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CCTGACCAACGAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTT 840

Qy        841 CAGGAGGCTCTTGAAAATCATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGGAGGCTCTTGAAAATCATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAG 900

Qy        901 TGACCCGATTGATATTCCAAACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TGACCCGATTGATATTCCAAACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCAT 960

Qy        961 CCTTCAGTATGAGAACAACATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCTTCAGTATGAGAACAACATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAG 1020

Qy       1021 CGAGACTGTCCTAGACTTCATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGAGACTGTCCTAGACTTCATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCA 1080

Qy       1081 GGAGCAAGTGTCGAAGGAGCTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGAGCAAGTGTCGAAGGAGCTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGAC 1140

Qy       1141 CTACCGGGTGGATGACATTGACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CTACCGGGTGGATGACATTGACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGA 1200

Qy       1201 TGGCTCGGAGGTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TGGCTCGGAGGTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGA 1260

Qy       1261 CCTGAAGCAGCCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CCTGAAGCAGCCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCT 1320

Qy       1321 GCCTGGCCCAGCTATGNTCATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAAT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GCCTGGCCCAGCTATGNTCATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAAT 1380

Qy       1381 GCGCAATGATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GCGCAATGATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCA 1440

Qy       1441 GCGGCAGCGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GCGGCAGCGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAG 1500

Qy       1501 AGAGCTTCGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAAT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AGAGCTTCGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAAT 1560

Qy       1561 CTTACAATCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTTACAATCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGC 1620

Qy       1621 AGACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AGACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTG 1680

Qy       1681 GCTGCTGATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTT 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GCTGCTGATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTT 1740

Qy       1741 CAAAGTGACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CAAAGTGACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGA 1800
```

```
Qy    1801 CAGAACAGAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGT 1860
Db    1801 CAGAACAGAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGT 1860

Qy    1861 TGTCTGTCTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTG 1920
Db    1861 TGTCTGTCTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTG 1920

Qy    1921 TACAGACTGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAAC 1980
Db    1921 TACAGACTGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAAC 1980

Qy    1981 AGTCATGGCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCT 2040
Db    1981 AGTCATGGCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCT 2040

Qy    2041 CTGGCGGGTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGA 2100
Db    2041 CTGGCGGGTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGA 2100

Qy    2101 CACCACAGCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGAC 2160
Db    2101 CACCACAGCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGAC 2160

Qy    2161 CCGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAA 2220
Db    2161 CCGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAA 2220

Qy    2221 GGTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCG 2280
Db    2221 GGTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCG 2280

Qy    2281 TGTCATCGTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGA 2340
Db    2281 TGTCATCGTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGA 2340

Qy    2341 GGTCCCACAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGAC 2400
Db    2341 GGTCCCACAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGAC 2400

Qy    2401 TGTAATCGTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACT 2460
Db    2401 TGTAATCGTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACT 2460

Qy    2461 TCAGAACCCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGA 2520
Db    2461 TCAGAACCCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGA 2520

Qy    2521 CTTTTTCATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGT 2580
Db    2521 CTTTTTCATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGT 2580

Qy    2581 CATCTATGACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCTG 2640
Db    2581 CATCTATGACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCTG 2640

Qy    2641 CCACGTGTACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACA 2700
Db    2641 CCACGTGTACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACA 2700

Qy    2701 CAAGCTGGCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAA 2760
Db    2701 CAAGCTGGCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAA 2760

Qy    2761 CCGCCTCTACTACCTCTAACCAGCAGCGAGGGCCGGCATGCTGGTTTCCTTCAGAGCTCT 2820
Db    2761 CCGCCTCTACTACCTCTAACCAGCAGCGAGGGCCGGCATGCTGGTTTCCTTCAGAGCTCT 2820

Qy    2821 GCCCAATCCTTCAGTTCTATTTTTCATAGGATGAGGTTCTAGCGGGATGTTGGGGGTAGG 2880
Db    2821 GCCCAATCCTTCAGTTCTATTTTTCATAGGATGAGGTTCTAGCGGGATGTTGGGGGTAGG 2880

Qy    2881 GGTTGGGGGCCGCTGTTTTGTTTTGTTTTGTTTTTAGTTGTTGGTGTGTAGCGGCTATAG 2940
Db    2881 GGTTGGGGGCCGCTGTTTTGTTTTGTTTTGTTTTTAGTTGTTGGTGTGTAGCGGCTATAG 2940

Qy    2941 AGTTAACTAAGGCTTATGTTTGCTTTGTTTCTCATGGTGTGTTGTGACAGTTTGCTTGGC 3000
Db    2941 AGTTAACTAAGGCTTATGTTTGCTTTGTTTCTCATGGTGTGTTGTGACAGTTTGCTTGGC 3000

Qy    3001 AGAGTGGTGGCAAGGGCCGTGGTACCACAGATAGACTTTCCTAAAGCAGCCGTGTTGGCA 3060
Db    3001 AGAGTGGTGGCAAGGGCCGTGGTACCACAGATAGACTTTCCTAAAGCAGCCGTGTTGGCA 3060

Qy    3061 GGTTCTTGGAGGTGCTCCTGGGAGCACCTCCTATTGTACGCTGCTCTCTCTCTCTCTCTC 3120
Db    3061 GGTTCTTGGAGGTGCTCCTGGGAGCACCTCCTATTGTACGCTGCTCTCTCTCTCTCTCTC 3120

Qy    3121 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGCATGTTTGAATTTAAAAGAGGA 3180
Db    3121 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGCATGTTTGAATTTAAAAGAGGA 3180

Qy    3181 AAGGACACAAGGCAAGAATGCCCAGAGCATCCTAGATGGGTGCTATATGGGACTCTGACA 3240
Db    3181 AAGGACACAAGGCAAGAATGCCCAGAGCATCCTAGATGGGTGCTATATGGGACTCTGACA 3240

Qy    3241 CCTCAGGTTCTTAAAGATAAAAAGGCCAGAGGTACCTTTCTGTAATTCTTTTAAGAGTTA 3300
Db    3241 CCTCAGGTTCTTAAAGATAAAAAGGCCAGAGGTACCTTTCTGTAATTCTTTTAAGAGTTA 3300

Qy    3301 GGTCATAAGGAATGAAAAGAGTTACTTTTCTCCCTCCATGAAAGTGTTTTTTAAGACTGA 3360
Db    3301 GGTCATAAGGAATGAAAAGAGTTACTTTTCTCCCTCCATGAAAGTGTTTTTTAAGACTGA 3360

Qy    3361 CAAGGCCGGTGGTGGAGTGAGTGTTGTTGTTTTGTTTCGCAATTTTCTTACCAAGTGAAA 3420
Db    3361 CAAGGCCGGTGGTGGAGTGAGTGTTGTTGTTTTGTTTCGCAATTTTCTTACCAAGTGAAA 3420

Qy    3421 TTACTGACCATAATTTTATTTAATAGTTTTTAGTTTTTGAAAGCCTGTGTTTTCTTTGCT 3480
Db    3421 TTACTGACCATAATTTTATTTAATAGTTTTTAGTTTTTGAAAGCCTGTGTTTTCTTTGCT 3480

Qy    3481 TTTAAACTACAGGCGACCACTCTGGGATAGGATTCCACTCCCAAAAGTAATGTGTTGGGT 3540
Db    3481 TTTAAACTACAGGCGACCACTCTGGGATAGGATTCCACTCCCAAAAGTAATGTGTTGGGT 3540

Qy    3541 GTGGTGGCACACGCCTGTGATCTCAGGACTCAGGAGGTGAAGCCAGGGGAACCCCAAGTT 3600
Db    3541 GTGGTGGCACACGCCTGTGATCTCAGGACTCAGGAGGTGAAGCCAGGGGAACCCCAAGTT 3600

Qy    3601 CCCGGTCAGCCTAGGCTACATAGTGAGACTTTGTCTTAACAAACAGAAAAAAAGAAAAAG 3660
Db    3601 CCCGGTCAGCCTAGGCTACATAGTGAGACTTTGTCTTAACAAACAGAAAAAAAGAAAAAG 3660

Qy    3661 AAAGAAAAAAAAAAAGAAAGAAAAAAAAAARAAAGAAAAATGGAGCAGTATGAGTTATGGGT 3720
Db    3661 AAAGAAAAAAAAAAAGAAAGAAAAAAAAAARAAAGAAAAATGGAGCAGTATGAGTTATGGGT 3720

Qy    3721 GCCAGTTTCTGCAGGAGACTTTGGGAGTAAGAGCTGCCTCTGTTCCTAAGAGGCTGCAAG 3780
Db    3721 GCCAGTTTCTGCAGGAGACTTTGGGAGTAAGAGCTGCCTCTGTTCCTAAGAGGCTGCAAG 3780

Qy    3781 GCTGCGGCTGGGGGGCCAGTTCAGGTGCCAGGTGAGGTACGGCCACATCTATTTGCCTGG 3840
Db    3781 GCTGCGGCTGGGGGGCCAGTTCAGGTGCCAGGTGAGGTACGGCCACATCTATTTGCCTGG 3840

Qy    3841 CTCTTTGACAAAGAGAGCAGCTTAGAGTGGCTAGAGTGTGATGAGTTTTACACAATGGCG 3900
Db    3841 CTCTTTGACAAAGAGAGCAGCTTAGAGTGGCTAGAGTGTGATGAGTTTTACACAATGGCG 3900

Qy    3901 ATGTGTGTGTGTGCCCCTGGCACAGCTCAATAAAGTTTCTGTCTCACTTTAAAAAAAAAA 3960
Db    3901 ATGTGTGTGTGTGCCCCTGGCACAGCTCAATAAAGTTTCTGTCTCACTTTAAAAAAAAAA 3960

Qy    3961 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4020
```

```
Db       3961 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4020

Qy       4021 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4064
Db       4021 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4064


RESULT 2
AAA07588
ID   AAA07588 standard; DNA; 3472 BP.
XX
AC   AAA07588;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Human piwi gene, designated hiwi.
XX
KW   Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW   anaemia; immunodeficiency; male infertility; human; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             167..2752
FT                   /*tag= a
FT                   /product= hiwi
FT                   /transl_except= (pos:392..394; aa:Xaa)
FT                   /transl_except= (pos:1073..1075; aa:Xaa)
FT                   /transl_except= (pos:2369..2370; aa:Xaa)
FT                   /note= "Xaa= Leu or Ile"
XX
PN   WO200032039-A1.
XX
PD   08-JUN-2000.
XX
PF   03-DEC-1999;   99WO-US28764.
XX
PR   04-DEC-1998;   98US-0110901.
XX
PA   (UYDU-) UNIV DUKE.
XX
PI   Lin H;
XX
DR   WPI; 2000-412085/35.
DR   P-PSDB; AAY90235.
XX
PT   Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT   therapy of diseases such as cancer and in various research and
PT   diagnostic applications -
XX
PS   Claim 19; Page 189-194; 201pp; English.
XX
CC   This sequence encodes the human piwi family protein, designated
CC   hiwi. The piwi family nucleic acids and polypeptides are used in gene
CC   therapy of diseases such as cancer and also in various research and
CC   diagnostic applications. The sequences can also be used to treat
CC   tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ   Sequence 3472 BP; 1044 A; 694 C; 796 G; 935 T; 3 other;

  Query Match             45.1%;  Score 1833.6;  DB 21;  Length 3472;
  Best Local Similarity   81.8%;  Pred. No. 1.1e-269;
  Matches 2125;  Conservative    0;  Mismatches  471;  Indels    3;  Gaps    1;

Qy        188 AAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGAGACGGTG 247
Db        164 ACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCG 223

Qy        248 CAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACCTCAACAG 307
Db        224 CAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCG 283

Qy        308 TCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGTCGGAGCC 367
Db        284 CCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAA---CAGCAGGAGGA 340

Qy        368 ACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACTGGCAGAG 427
Db        341 ACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAG 400

Qy        428 AGAGGAGGGCGTCGCCGAGACTTCCATGACNTTGGTGTGAACACCAGACAGAACCTTGAC 487
Db        401 AGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGAC 460

Qy        488 CATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCAC 547
Db        461 CATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCAT 520

Qy        548 TTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCC 607
Db        521 TTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCA 580

Qy        608 CTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGA 667
Db        581 CTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGA 640

Qy        668 AGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGTC 727
Db        641 AAGTGTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTT 700

Qy        728 ACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCACCCTGACC 787
Db        701 ACTGAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACA 760

Qy        788 AACGAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTTCAGGAGG 847
Db        761 AATGAACTTCCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGG 820

Qy        848 CTCTTGAAAATCATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAGTGACCCG 907
Db        821 CTTTTGAAAATCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCA 880

Qy        908 ATTGATATTCCAAACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCATCCTTCAG 967
Db        881 ATTGATATTCCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAG 940

Qy        968 TATGAGAACAACATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAGCGAGACT 1027
Db        941 TATGAAAACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACT 1000

Qy       1028 GTCCTAGACTTCATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCAGGAGCAA 1087
Db       1001 GTTTTGGATTTCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAA 1060

Qy       1088 GTGTCGAAGGAGCTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGACCTACCGG 1147
Db       1061 GTTTCCAAAGAANTAATAGGTTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGA 1120

Qy       1148 GTGGATGACATTGACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGATGGCTCG 1207
Db       1121 GTGGATGATATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCT 1180

Qy       1208 GAGGTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGACCTGAAG 1267
Db       1181 GGGGTCAGCTTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAG 1240

Qy       1268 CAGCCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCTGCCTGGC 1327
Db       1241 CAGCCTGTCTTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGG 1300

Qy       1328 CCAGCTATGNTCATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAATGCGCAAT 1387
Db       1301 CCTGCCATGCTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAAT 1360

Qy       1388 GATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCAGCGGCAG 1447
```

```
Db       1361 GATTTTAACGTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAG 1420
Qy       1448 CGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAGAGAGCTT 1507
Db       1421 CGTGAAGTGGGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTT 1480
Qy       1508 CGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAATCTTACAA 1567
Db       1481 CGAGACTGGGGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAA 1540
Qy       1568 TCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGCAGACTGG 1627
Db       1541 ACAGAAAAGATTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGG 1600
Qy       1628 TCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGGCTGCTG 1687
Db       1601 TCCAAAGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTG 1660
Qy       1688 ATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTTCAAAGTG 1747
Db       1661 ATCTATACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTT 1720
Qy       1748 ACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGACAGAACA 1807
Db       1721 ACACCAGCCATGGGCATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACT 1780
Qy       1808 GAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGT 1867
Db       1781 GAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGT 1840
Qy       1868 CTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTGTACAGAC 1927
Db       1841 CTGTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGAT 1900
Qy       1928 TGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATG 1987
Db       1901 TGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATG 1960
Qy       1988 GCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGG 2047
Db       1961 GCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGG 2020
Qy       2048 GTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACACCACA 2107
Db       2021 GTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACA 2080
Qy       2108 GCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGG 2167
Db       2081 GCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGG 2140
Qy       2168 TTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGC 2227
Db       2141 TTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGC 2200
Qy       2228 TTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATC 2287
Db       2201 CTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATC 2260
Qy       2288 GTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCA 2347
Db       2261 GTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCA 2320
Qy       2348 CAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATC 2407
Db       2321 CAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATT 2380
Qy       2408 GTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAAC 2467
Db       2381 GTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAAT 2440
Qy       2468 CCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTC 2527
Db       2441 CCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTT 2500
Qy       2528 ATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTAT 2587
Db       2501 ATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTAT 2560
Qy       2588 GACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTG 2647
Db       2561 GACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATC 2620
Qy       2648 TACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTG 2707
Db       2621 TATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTG 2680
Qy       2708 GCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTC 2767
Db       2681 GCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTT 2740
Qy       2768 TACTACCTCTAACCAGCAG 2786
Db       2741 TACTACCTCTAACCTGCAG 2759
```

RESULT 3
AAS13630
ID AAS13630 standard; cDNA; 3649 BP.
XX
AC AAS13630;
XX
DT 18-DEC-2001 (first entry)
XX
DE cDNA sequence encoding mammalian Spg16.
XX
KW Mammalian; reproductive-specific protein; male infertility; gene therapy;
KW spermatogenesis; sperm count disorder; anti infertility;
KW reproduction; ss.
XX
OS Mammalia.
XX
PN WO200166752-A2.
XX
PD 13-SEP-2001.
XX
PF 07-MAR-2001; 2001WO-US07371.
XX
PR 07-MAR-2000; 2000US-0187518.
PR 12-JAN-2001; 2001US-0261557.
XX
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
XX
PI Wang PJ, Page DC;
XX
DR WPI; 2001-570774/64.
DR P-PSDB; AAU07866.
XX
PT Novel reproduction-specific protein, useful for treating disorders of
PT reduced sperm count, enhancing/increasing sperm count and/or sperm
PT activity -
XX
PS Claim 1; Fig 15; 151pp; English.
XX
CC The present invention relates to the isolation of novel mammalian and
CC human reproductive-specific proteins (AAU07859-AAU07899), and the
CC nucleic acids encoding them. The nucleic acids encoding
CC reproductive-specific proteins are useful for diagnosing infertility
CC which is a result of reduced sperm count, reduced sperm motility,
CC malformed sperm or combinations of these. The sequences of the invention
CC are useful as markers for spermatogonial cells, for identifying genes or
CC proteins characteristic of male infertility, diagnosing or aiding in
CC the diagnosis of infertility in men, and for contraception in which
CC sperm production or sperm count is reduced or defective sperm is
CC produced. Antibodies to reproductive-specific proteins are useful for
CC determining the presence of these proteins in a sample obtained from a

CC man being assessed for infertility, for identifying the expression of
CC genes in particular cell type or particular developmental stage, for
CC studies of spermatogenesis, and for immunofluorescence of germ cells or
CC in Western blots for assessing the presence of the protein the antibody
CC binds. The sequences of the invention are also useful for treating
CC disorders of reduced sperm count, and for increasing sperm count and/or
CC sperm activity. The nucleic acids of the invention are useful in gene
CC therapy. AAS13623-AAS13647 represent cDNA sequences encoding for the
CC mammalian reproduction-specific proteins of the present invention.
XX
SQ Sequence 3649 BP; 947 A; 841 C; 911 G; 950 T; 0 other;

Query Match 9.8%; Score 399.4; DB 22; Length 3649;
Best Local Similarity 53.1%; Pred. No. 5.7e-52;
Matches 991; Conservative 0; Mismatches 853; Indels 21; Gaps 6;

```
Qy        941 CCCGGCTTCACCACCTCCATCCTTCAGTATGAGAACAACATCATGCTCTGCACAGACGTC 1000
              || ||||   |  |    ||||    |   ||        || | ||   | | || |||
Db          1 CCTGGCTATGCGGCTAGTATCCGGAGGACAGACGGGGGTCTCTTCCTGCTCGCTGATGTC 60

Qy       1001 AGCCACAAGGTGCTCCGCAGCGAGACTGTCCTAGACTTCATGTTCAATCTATACCAGCAG 1060
              || |||||  | || | |||  |||| || ||  |||||     | | |||||||||
Db         61 TCTCATAAGGTCATTCGGAACGACTCTGTGCTGGATGTCATGCATGCTATCTACCAGCAG 120

Qy       1061 ACAGAGGAGCACAAGTTCCAGGAGCAAGTGTCGAAGGAGCTCATAGGCCTCATCGTTCTC 1120
              |  ||||||||   ||||||||  |        |||   ||  | |||  ||| ||  ||
Db         121 AACAAGGAGCAC---TTCCAGGACGAGTGCAGCAAGCTTCTGGTTGGCAGCATTGTCATC 177

Qy       1121 ACCAAGTACAATAACAAGACCTACCGGGTGGATGACATTGACTGGGACCAGAATCCAAAG 1180
              ||    |||| ||    ||||||||  | |||||  | ||||||  || |||  || ||
Db         178 ACGCGCTACAACAATCGTACCTACCGAATCGATGATGTGGACTGGAACAAGACCCCTAAA 237

Qy       1181 AGCACCTTCAAGAAGGCGGATGGCTCGGAGGTCAGCTTCCTGGAGTACTACAGGAAGCAA 1240
              || |||    | | |||| ||   |||  |||  |||||||| ||||||||  ||  |
Db         238 GACAGCTTTGTCATGTCGGACGGGAAGGAAATCACATTCCTGGAATACTACAGCAAAAAC 297

Qy       1241 TACAACCAGGAGATCACGGACCTGAAGCAGCCGGTGCTGGTGAGCCAACCCA---AGCGG 1297
              ||          ||| |||     | ||||||  |||||  |   ||  ||||   || |
Db         298 TATGGGATCACAGTCAAGGAAGATGACCAGCCGCTGCTGATCCACCGGCCCAGTGAGAGA 357

Qy       1298 AGGAGAGGCCCCGGCGGCACCCTGCCTGGCCCAGCTATGNTCATCCCTGAACTCTGCTAT 1357
              ||   ||  |||     |||  |||        || |  | || || |||| ||
Db         358 CAGAATAACCATGGCATGTTGCTGAAGGGCGAGATCCTGCTGCTGCCCGAGCTCTCCTTC 417

Qy       1358 CTCACAGGCCTGACTGATAAAATGCGCAATGATTTCAATGTGATGAAGGACCTGGCAGTG 1417
              | || ||  |  |||| || |||   || || ||||  |  |||||||| || |   |
Db         418 ATGACGGGGATCCCTGAGAAGATGAAGAAGGACTTCAGGGCCATGAAGGACTTGACTCAG 477

Qy       1418 CACACGCGGCTGACCCCTGAGCAGCGGCAGCGGGAGGTGGGCCGCCTCATCGACTACATC 1477
              || |     |||| |||  ||||||  ||  | |   |||   |||| |  |    ||
Db         478 CAGATTAACCTGAGCCCCAAGCAGCACCACGGTGCTTTGGAATGCCTGCTGCAGAGAATT 537

Qy       1478 CACAAGGATGACAATGTGCAGAGAGAGCTTCGAGACTGGGGCCTGAGCTTCGACTCAAAC 1537
              |  | ||  |  |     |  |||||      |||||| || ||  |  |   |  |
Db         538 TCACAAAACGAGACAGCCAGCAATGAGCTGACCCGCTGGGGGCTCAGTCTGCATAAAGAT 597

Qy       1538 TTGCTGTCCTTCTCTGGAAGAATCTTACAATCTGAGAAGATCCACCAGGGCGGAAAGACG 1597
              | |      |   ||  |  | | |  |||| |||| ||        | || ||
Db         598 GTCCACAAGATTGAAGGTCGGCTTCTGCCAATGGAGAGGATCAAC---TTAAGGAACACT 654

Qy       1598 TTTGATTACAACCCACAATTTGCAGACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGATC 1657
              |   |  ||   |  |          |||||   || |||   |   | | |   ||
Db         655 TCATTTGTCACATCGGAGGGCCTGAACTGGGTTAAGGAAGTGACCAGAGATGCTTCCATT 714

Qy       1658 AGCGTGAAGCCATTGGATAACTGGCTGCTGATCTATACCCGCAGGAATTATGAAGCAGCC 1717
              .  ||  || ||  ||||   ||  | ||| |    ||        ||   ||||
Db         715 CTAACTATTCCCATGCATTTCTGGGCACTCTTTTATCCAAAGAGAGCAATGGACCAAGCC 774

Qy       1718 AACTCACTGATACAGAACCTGTTCAAAGTGACTCCAGCCATGGGCATCCAGATGAAAAAG 1777
              |    |||| |  | |   ||   ||  | |      |||| |||||  |  |  |
Db         775 AGAGAACTGGTTAACATGTTGGAAAAGATTGCCGGGCCCATTGGCATGCGCACAAGCCCC 834

Qy       1778 GCAATCATGATCGAGGT---GGATGACAGAACAGAAGCTTATCTGAGAGCCTTGCAG--- 1831
```

```
Db        835 CCAGCCTGGGTTGAGCTGAAGGATGACCGAATAGAGACCTATATCAGGACCATTCAGTCC 894

Qy       1832 ---CAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGTCTCTTGTCAAGTAATCGGAAG 1888
                | |   ||   |   | |  || || || || ||  ||  |     | |  ||  |
Db         895 TTACTGGGAGTTGAGGGGAAGATACAAATGGTCGTTTGCATCATCATGGGCACACGTGAT 954

Qy       1889 GACAAATATGATGCCATCAAGAAGTACTTGTGTACAGACTGCCCCACCCCAAGTCAGTGT 1948
              ||    ||||  ||||||||||||    |  ||      | | |||    ||    |||
Db         955 GATCTCTATGGAGCCATCAAGAAGCTGTGCTGCGTGCAGTCCCCAGTGCCCTCACAGGTC 1014

Qy       1949 GTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATGGCCATTGCCACCAAGATCGCC 2008
              |    | ||| ||| | ||  |||      |   |   |  | | ||    || ||
Db        1015 ATCAATGTCCGAACCATTGGTCAGCCCACCAGGCTTCGGAGCGTGGCTCAGAAAATTTTA 1074

Qy       2009 CTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGGGTGGACATGGCCCTGAAACTG 2068
              || |||||||||| ||  |||| || |||||||| | |||||  ||  | |||||||
Db        1075 CTTCAGATGAACTGTAAACTGGGTGGTGAGCTCTGGGGAGTGGATATTCCGCTGAAACAA 1134

Qy       2069 GCAATGATCGTGGGCATCGACTGTTACCATGACACCACAGCTGGGCGGAGGTCCATCGCA 2128
              ||||  |  | || || ||    |||||||| |||     ||   | | ||  | |
Db        1135 CTAATGGTGATTGGAATGGATGTGTACCATGACCCCAGCAGAGGCATGCGCTCTGTGGTC 1194

Qy       2129 GGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCCCGCTGCGTCTTTCAG 2188
              || ||||| |||||||| |||        | |||   |||| ||| |    || || |||
Db        1195 GGCTTCGTGGCCAGCATAAATCTCACACTCACCAAATGGTACTCGAGGGTGGTGTTCCAG 1254

Qy       2189 GACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGCTTGCAAGCTGCTCTGAGGGCT 2248
                |  |||||| | |||||  | || ||| | ||| ||   | | |  |||
Db        1255 ATGCCACATCAGGAGATTGTGGACAGCCTGAAGCTCTGCCTGGTGGGTTCCTTGAAAAAG 1314

Qy       2249 TGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATCGTGTACCGAGACGGTGTGGGG 2308
              |    ||      ||  | |   | ||      |  | ||||||||||| || |||
Db        1315 TATTATGAGGTGAACCATTGTCTCCCAGAGAAAATTGTGGTGTACCGAGATGGAGTGTCT 1374

Qy       2309 GACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCACAGTTCCTAGATTGCCTCAAG 2368
              || || ||||| |||||  | | ||| || ||| |||| ||| | |   | ||  |  |
Db        1375 GATGGCCAGCTAAAGACAGTTGCCAACTACGAGATCCCTCAGCTGCAGAAGTGTTTTGAA 1434

Qy       2369 TCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATCGTGGTGAAGAAGCGTGTCAAT 2428
              |  | |  |    ||| |||| |   ||   ||  | || ||  |||||    |||
Db        1435 GCCTTTGATA---ACTACCACCCCAAGATGGTGGTGTTTGTAGTTCAGAAGAAAATCAGC 1491

Qy       2429 GCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAACCCTCTTCCAGGGACAGTCATC 2488
              |||   | |    |   ||  |      |  | |||    || |||||  ||  |
Db        1492 ACCAATCTGTACCTTGCTGCTCCTGATCACTTCGTAACCCCCTCCCCCGGGACTGTGGTT 1551

Qy       2489 GATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTCATCGTGAGCCAGGCAGTGAGA 2548
              |||       | |||||    ||||||   || || | | |   |||    ||| ||
Db        1552 GATCATACCATAACCAGCTGTGAGTGGGTGGATTTCTACCTTCTTGCCCATCATGTGCGA 1611

Qy       2549 AGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTATGACAGCAGTGGCCTGAAGCCC 2608
              ||  |||      || ||||||||||     |     |||       ||||  ||
Db        1612 CAGGGCTGTGGCATACCTACACACTACATCTGTGTTCTGAACACTGCAAATCTGAGCCCT 1671

Qy       2609 GACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTGTACTATAATTGGCCTGGAGTC 2668
              || ||||| ||| || |||| | |||  | |||||| ||||||  ||||||||||    |
Db        1672 GATCACATGCAGAGGTTGACTTTCAAACTATGCCACATGTACTGGAATTGGCCTGGTACC 1731

Qy       2669 ATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTGGCCTTCCTCGTGGGCCAGAGC 2728
              ||||||||  || || |||||| |||||||  |||||||  || ||||    || ||||
Db        1732 ATCCGAGTTCCAGCTCCTTGCAAGTATGCCCACAAGCTAGCTTTCCTGTCCGGACAGATT 1791

Qy       2729 ATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTCTACTACCTCTAACCAGCAGCG 2788
              | ||    ||||||  | ||   ||||        ||| | || ||| |||| | | |
Db        1792 TTGCATCATGAGCCAGCCATCCAGCTGTGTGGGAACCTGTTCTTCCTGTAACTGGGAACT 1851

Qy       2789 AGGGC 2793
              || |
Db        1852 TGGAC 1856
```

RESULT 4

```
AAH15960
ID   AAH15960 standard; cDNA; 2272 BP.
XX
AC   AAH15960;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:14558.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 14558; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2272 BP; 609 A; 509 C; 585 G; 569 T; 0 other;

  Query Match             8.5%;  Score 345;  DB 22;  Length 2272;
  Best Local Similarity   52.1%;  Pred. No. 1e-43;
  Matches  893;  Conservative    0;  Mismatches  802;  Indels   18;  Gaps    5;

Qy       1080 AGGAGCAAGTGTCGAAGGAGCTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGA 1139
             |||| |     | |||   ||  | |||   || ||| |||||   || || ||    |
Db          1 AGGATGAGTGTACTAAGCTTCTGGTTGGCAATATTGTTATCACCTGATATAACAATCGTA 60

Qy       1140 CCTACCGGGTGGATGACATTGACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGG 1199
```

```
Db         61 CCTATCGTATTGATGATGTGGATTGGAATAAGACTCCAAAGGATAGCTTCACGATGTCTG 120

Qy       1200 ATGGCTCGGAGGTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGG 1259
             ||||    ||| |||  ||| |||| |||||||| ||  | ||           | | ||
Db         121 ATGGGAAAGAGATCACATTCTTGGAATACTACAGCAAAAATTATGGGATCACAGTTAAGG 180

Qy       1260 ACCTGAAGCAGCCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCC 1319
             |   | | |||||  |||||  |   |  ||||   |||   |        || | |
Db         181 AAGAGGACCAGCCATTGCTGATTCACAGGCCCAGTGAGAGACAGGATAATCATGGGATGC 240

Qy       1320 TGC---CTGGCCCAGCTATGNTCATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATA 1376
             |||     ||   |    || |  | |||||.|| |  | | | || ||  |  | || |
Db         241 TGCTAAAAGGGGAAATCCTGCTGCTGCCTGAGCTTTCTTTTATGACCGGAATCCCAGAGA 300

Qy       1377 AAATGCGCAATGATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTG 1436
             | |||   || || ||||  |  |||||||  ||||   ||| |     |||| |||
Db         301 AGATGAAGAAGGACTTCAGAGCCATGAAGGATTTGGCTCAGCAAATCAATCTGAGCCCCA 360

Qy       1437 AGCAGCGGCAGCGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGC 1496
             |||| |  ||  | |   |||  || |  | |    ||    ||| | ||    |
Db         361 AGCAACACCATAGTGCTTTGGAATGCTTGCTGCAAAGAATTGCAAAGAACGAGGCAGCCA 420

Qy       1497 AGAGAGAGCTTCGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAA 1556
               | || ||      ||||| ||  |  |  |   | | |     |   |||
Db         421 CCAATGAACTGATGCGTTGGGGGCTCCGTCTGCAAAAGGATGTACATAAGATTGAAGGAC 480

Qy       1557 GAATCTTACAATCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAAT 1616
             |  |  | | |   || ||  ||  ||       ||| || |   |  || | ||
Db         481 GTGTTCTGCCAATGGAAAGAATTAACTTAAA---AAATACTTCGTTTATCACATCTCAGG 537

Qy       1617 TTGCAGACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATA 1676
               | |||||   || |||  ||   | |  ||    |||       |  ||  || ||
Db         538 AACTAAACTGGGTTAAGGAAGTAACCAGAGACCCTTCCATCTTGACTATCCCCATGCATT 597

Qy       1677 ACTGGCTGCTGATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACC 1736
               ||||  ||  | ||  |   ||       ||    ||      ||||  |  |
Db         598 TCTGGGCACTTTTTTACCCAAAGAGAGCAATGGACCAGGCTCGAGAACTGGTCAACATGT 657

Qy       1737 TGTTCAAAGTGACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGT-- 1794
             ||   ||  |  |      |||| ||||| |  ||||      |   |  | | ||  |
Db         658 TGGAGAAGATAGCCGGCCCCATTGGCATGCGTATGAGCCCACCGGCCTGGGTTGAACTAA 717

Qy       1795 -GGATGACAGAACAGAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAG------ 1847
              |||||||  ||| |||  ||||| | ||| || | ||     | | |      | |
Db         718 AGGATGACCGAATAGAGACTTATGTCAGAACCATTCAATCCACGTTAGGAGCTGAGGGGA 777

Qy       1848 ACACTCAGATAGTTGTCTGTCTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCA 1907
             | |  ||||| |||||  ||  | |      |  ||  |  ||||  |||||||
Db         778 AGATACAGATGGTTGTTTGCATCATCATGGGCCCACGTGATGATCTCTATGGGGCCATCA 837

Qy       1908 AGAAGTACTTGTGTACAGACTGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGG 1967
             |||||  |  |||   | | |||    ||    |||  |||    |  || ||| | |
Db         838 AGAAGCTGTGCTGTGTGCAGTCCCCAGTGCCCTCCCAGGTTGTCAATGTTCGAACCATTG 897

Qy       1968 GCAAGCAGCAAACAGTCATGGCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGA 2027
             |  |||      |   |   |  | |||   |||||    || ||||| ||||| ||
Db         898 GTCAGCCCACCAGGCTTCGGAGTGTGGCCCAGAAGATTTTACTTCAGATTAACTGTAAAT 957

Qy       2028 TGGGAGGCGAGCTCTGGCGGGTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCG 2087
             |||| || |||||||||  | ||||| ||  |  |||||||  |  |||| |  | || || |
Db         958 TGGGTGGTGAGCTCTGGGGAGTGGATATTCCTCTGAAACAGTTAATGGTGATCGGGATGG 1017

Qy       2088 ACTGTTACCATGACACCACAGCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCA 2147
             |   ||||||||||  |||     ||   | | ||| | |  || |||||  || |||||||
Db         1018 ATGTTTACCATGACCCCAGTAGAGGCATGCGCTCCGTGGTTGGCTTCGTGGCAAGCATCA 1077

Qy       2148 ATGAAGGGATGACCCGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGG 2207
             ||        | ||    |||| |||||    || || |||   |     ||||||  | |
Db         1078 ATCTCACCCTCACAAAATGGTATTCCCGGGTGGTGTTCCAGATGCCGCATCAGGAGATTG 1137

Qy       2208 TGGATGGTCTCAAGGTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAAT 2267
             ||||  | || ||| | ||| |     |   | | |      |    ||      ||  | |
```

```
Db       1138 TGGACAGCCTGAAGCTATGCCTCGTGGGCTCCTTAAAAAAGTTTTATGAGGTGAACCACT 1197

Qy       2268 ACATGCCCAGCCGTGTCATCGTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCC 2327
Db       1198 GTCTACCAGAGAAGATTGTGGTGTACCGTGATGGAGTGTCTGATGGCCAACTGAAGACAG 1257

Qy       2328 TGGTCAATTATGAGGTCCCACAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACA 2387
Db       1258 TTGCCAACTATGAGATTCCTCAACTACAGAAGTGTTTTGAAGCTTTTGAGA---ATTATC 1314

Qy       2388 ACCCAAGACTGACTGTAATCGTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGT 2447
Db       1315 AGCCCAAGATGGTGGTGTTTGTAGTTCAGAAGAAAATCAGTACTAATCTATATCTGGCTG 1374

Qy       2448 CTGGGGGAAGACTTCAGAACCCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGAC 2507
Db       1375 CTCCTCAGAACTTTGTAACTCCCACTCCTGGAACTGTGGTAGATCATACAATAACAAGCT 1434

Qy       2508 CAGAGTGGTATGACTTTTTCATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAA 2567
Db       1435 GTGAGTGGGTGGATTTCTATCTTCTTGCCCATCATGTACGGCAGGGCTGTGGCATTCCTA 1494

Qy       2568 CACACTACAATGTCATCTATGACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGA 2627
Db       1495 CGCATTATGTCTGTGTTCTCAACACCGCAAACCTGAGCCCTGATCATATGCAGAGGCTGA 1554

Qy       2628 CATACAAGNTCTGCCACGTGTACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTT 2687
Db       1555 CTTTCAAACTGTGCCACATGTACTGGAATTGGCCTGGCACCATCAGAGTTCCAGCTCCTT 1614

Qy       2688 GCCAGTATGCACACAAGCTGGCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACC 2747
Db       1615 GCAAGTATGCCCACAAGCTAGCTTTCCTGTCAGGACACATCTTGCATCATGGACCAGCCA 1674

Qy       2748 TCTCCCTGTCCAACCGCCTCTACTACCTCTAAC 2780
Db       1675 TCCAGCTGTGCGAGAACCTGTTCTTCCTGTGAC 1707
```

RESULT 5
ABL10535
ID ABL10535 standard; cDNA; 3524 BP.
XX
AC ABL10535;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 26087.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB66432.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 26087; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3524 BP; 1053 A; 741 C; 761 G; 969 T; 0 other;

Query Match 7.8%; Score 318.2; DB 23; Length 3524;
Best Local Similarity 48.9%; Pred. No. 1.2e-39;
Matches 1120; Conservative 0; Mismatches 1141; Indels 29; Gaps 9;

```
Qy        500 TCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCACTTCCGGCTGACC 559
Db        345 TCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTCCGATTAAAA 404

Qy        560 TCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCCCTGATGGAGGCC 619
Db        405 ACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGAGCCGAGCATCGAGAAT 464

Qy        620 CGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGAAGGTGTCATGCT 679
Db        465 CCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTATCTA 524

Qy        680 TTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGTCACAGAAGTATTC 739
Db        525 TTCGATGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTCAGC 584

Qy        740 AGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCACCCTGACCAACGAGCTGCCG 799
Db        585 GGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATATCG 644

Qy        800 CCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTTCAGGAGGCTCTTGAAAATC 859
Db        645 TGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCCATGAAGGGC 704

Qy        860 ATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAGTGACCCGATTGATATTCCA 919
Db        705 CTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATAAGG 764

Qy        920 AACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCATCCTTCAGTATGAGAACAAC 979
Db        765 GAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAAGAT 824

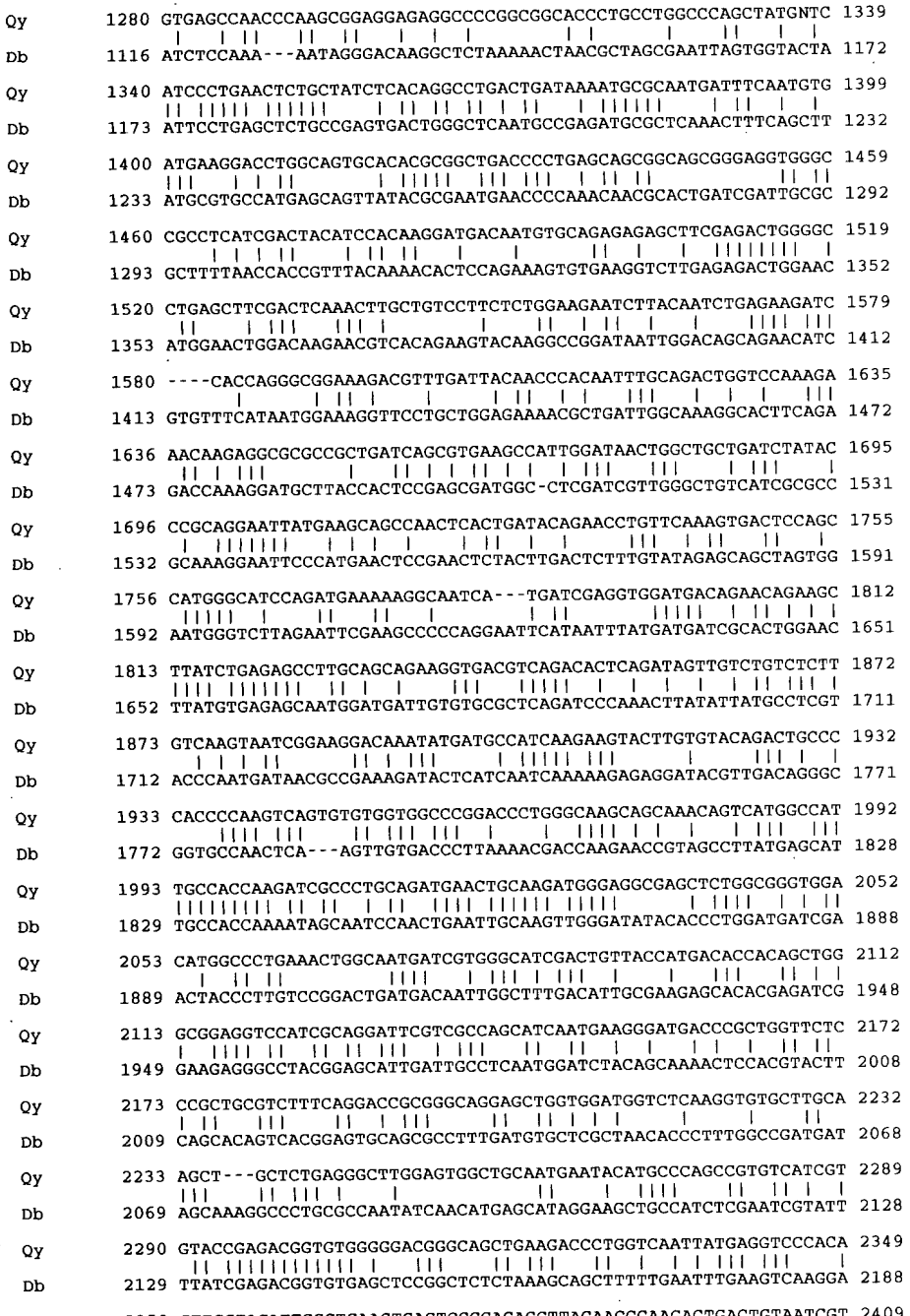
Qy        980 ATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAGCGAGACTGTCCTAGACTTC 1039
Db        825 ATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGACATA 884

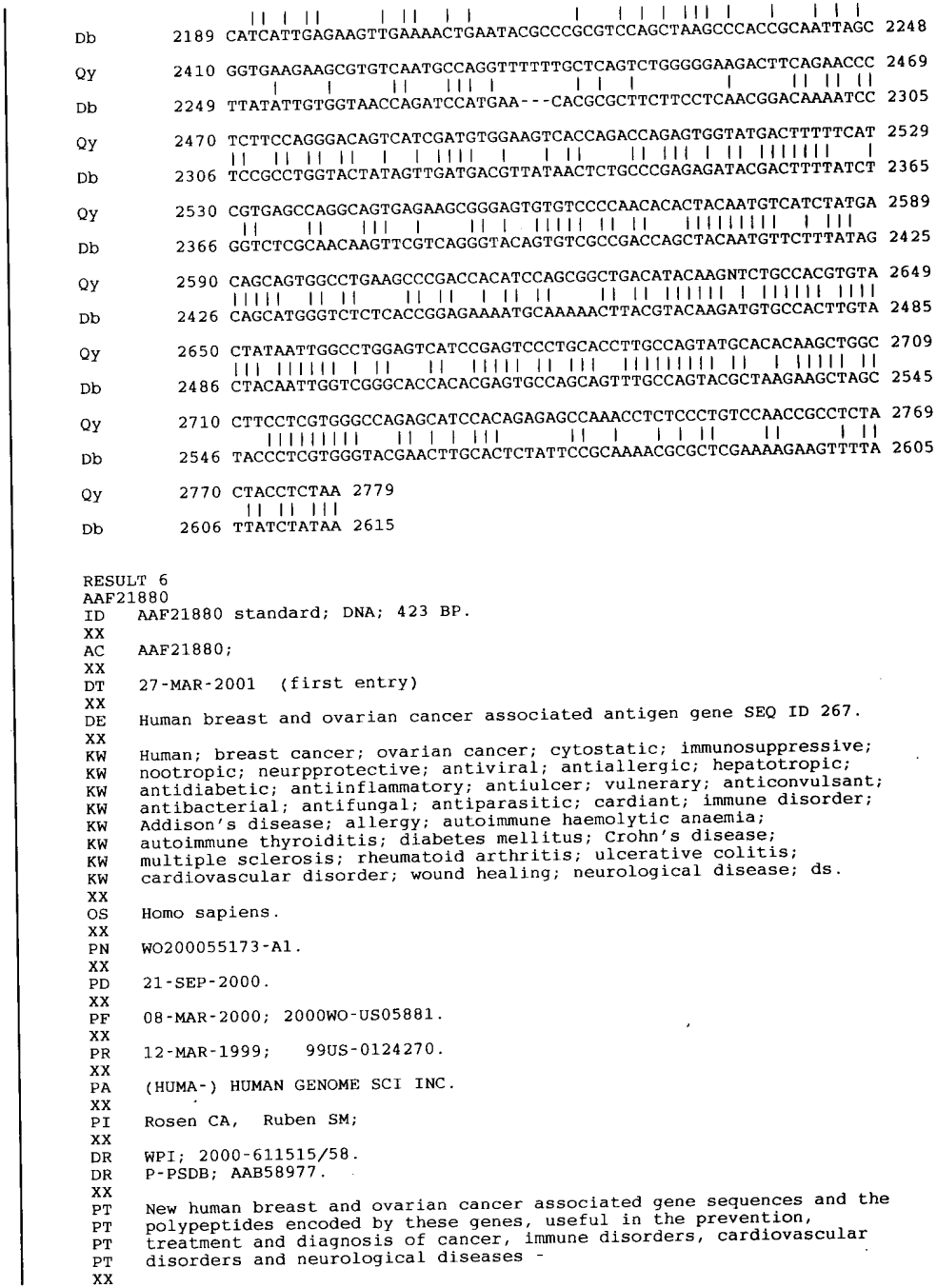
Qy       1040 ATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCAGGAGCAAGTGTCGAAGGAG 1099
Db        885 ATGCGACGTTGCTCACACAATCCGGCTCGTCA------TCAGGACGAAGTACGGGTAAAT 938

Qy       1100 CTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGACCTACCGGGTGGATGACATT 1159
Db        939 GTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGATGTC 998

Qy       1160 GACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGATGGCTCGGAGGTCAGCTTC 1219
Db        999 GACTTTGGACAAACTCCGAAATCAACATTCAGTTGCAAGGGTAG---AGATATCAGTTTC 1055

Qy       1220 CTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGACCTGAAGCAGCCGGTGCTG 1279
Db       1056 GTGGAATACTATCTCACTAAATATAATATACGCATTCGCGACCACAATCAGCCGCTGCTG 1115
```





RESULT 6
AAF21880
ID AAF21880 standard; DNA; 423 BP.
XX
AC AAF21880;
XX
DT 27-MAR-2001 (first entry)
XX
DE Human breast and ovarian cancer associated antigen gene SEQ ID 267.
XX
KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW nootropic; neurpprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW Addison's disease; allergy; autoimmune haemolytic anaemia;
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW cardiovascular disorder; wound healing; neurological disease; ds.
XX
OS Homo sapiens.
XX
PN WO200055173-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05881.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-611515/58.
DR P-PSDB; AAB58977.
XX
PT New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases -
XX

```
PS   Claim 1; Page 686; 1299pp; English.
XX
CC   Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC   proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC   associated with breast and ovarian cancer. Included in the invention are
CC   sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC   isolation and characterisation of the DNA and protein sequences of the
CC   invention. The breast and ovarian cancer associated DNA, protein, agonist
CC   or antagonist sequences exhibit cytostatic; immunosuppressive;
CC   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
CC   antibacterial; antifungal; antiparasitic and cardiant activity. The
CC   polynucleotide and protein sequences are used in the diagnosis of cancer,
CC   particularly breast and ovarian cancer. The nucleic acid sequences,
CC   proteins, agonists and agonists may also be used in the diagnosis,
CC   prevention and treatment of immune disorders e.g. Addison's disease,
CC   allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC   diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC   arthritis and ulcerative colitis; cardiovascular disorders such as
CC   myocardial ischaemias; wound healing; neurological diseases such as
CC   cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ   Sequence 423 BP; 104 A; 98 C; 129 G; 88 T; 4 other;

  Query Match             7.7%;  Score 311.8;  DB 21;  Length 423;
  Best Local Similarity   86.1%;  Pred. No. 1.1e-38;
  Matches  354;  Conservative    2;  Mismatches   54;  Indels    1;  Gaps    1;

Qy       1925 GACTGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTC 1984
Db         14 GATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTC 73

Qy       1985 ATGGCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGG 2044
Db         74 ATGGCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGRGAGCTCTGG 133

Qy       2045 CGGGTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACACC 2104
Db        134 AGGGTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTKACCATGACATG 193

Qy       2105 ACAGCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGC 2164
Db        194 ACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGC 253

Qy       2165 TGGTTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTG 2224
Db        254 TGGTTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTC 313

Qy       2225 TGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTC 2284
Db        314 TGCCTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATC 373

Qy       2285 ATCGTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAAT 2335
Db        374 ATCGTGTACCGSG-TGGCGTAGGAGACGGCCAGYTGAAAACACTGGTGACT 423


RESULT 7
AAA07586
ID   AAA07586 standard; DNA; 3047 BP.
XX
AC   AAA07586;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Drosophila piwi gene.
XX
KW   Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW   anaemia; immunodeficiency; male infertility; Drosophila; ds.
XX
OS   Drosophila sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             84..2615
FT                   /*tag= a
FT                   /product= piwi
FT                   /transl_except= (pos:120..122; aa:Xaa)
FT                   /transl_except= (pos:399..401; aa:Xaa)
FT                   /transl_except= (pos:2436..2438; aa:Xaa)
FT                   /note= "Xaa= Leu or Ile"
XX
PN   WO200032039-A1.
XX
PD   08-JUN-2000.
XX
PF   03-DEC-1999;   99WO-US28764.
XX
PR   04-DEC-1998;   98US-0110901.
XX
PA   (UYDU-) UNIV DUKE.
XX
PI   Lin H;
XX
DR   WPI; 2000-412085/35.
DR   P-PSDB; AAY90233.
XX
PT   Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT   therapy of diseases such as cancer and in various research and
PT   diagnostic applications -
XX
PS   Claim 19; Page 171-175; 201pp; English.
XX
CC   This sequence encodes the Drosophila piwi family protein, designated
CC   piwi. The piwi family nucleic acids and polypeptides are used in gene
CC   therapy of diseases such as cancer and also in various research and
CC   diagnostic applications. The sequences can also be used to treat
CC   tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ   Sequence 3047 BP; 936 A; 664 C; 672 G; 772 T; 3 other;

  Query Match             7.7%;  Score 311.4;  DB 21;  Length 3047;
  Best Local Similarity   48.7%;  Pred. No. 1.3e-38;
  Matches 1115;  Conservative    0;  Mismatches 1146;  Indels   29;  Gaps    9;

Qy        500 TCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCACTTCCGGCTGACC 559
Db        345 TCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGANTAAAA 404

Qy        560 TCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCCCTGATGGAGGCC 619
Db        405 ACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAGAAT 464

Qy        620 CGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGAAGGTGTCATGCT 679
Db        465 CCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTATCTA 524

Qy        680 TTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGTCACAGAAGTATTC 739
Db        525 TTCGACGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTCAGC 584

Qy        740 AGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCACCCTGACCAACGAGCTGCCG 799
Db        585 GGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATATCG 644

Qy        800 CCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTTCAGGAGGCTCTTGAAAATC 859
Db        645 TGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCGATGAAGGGC 704

Qy        860 ATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAGTGACCCGATTGATATTCCA 919
Db        705 CTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATAAGG 764

Qy        920 AACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCATCCTTCAGTATGAGAACAAC 979
Db        765 GAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAAGAT 824
```

```
Qy       980 ATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAGCGAGACTGTCCTAGACTTC 1039
Db       825 ATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGACATA 884

Qy      1040 ATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCAGGAGCAAGTGTCGAAGGAG 1099
Db       885 ATGCGACGTTGCTCACACAATCCGGCTCGTCA------TCAGGACGAAGTACGGGTAAAT 938

Qy      1100 CTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGACCTACCGGGTGGATGACATT 1159
Db       939 GTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGATGTC 998

Qy      1160 GACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGATGGCTCGGAGGTCAGCTTC 1219
Db       999 GACTTTGGACAAACTCCGAAATCAACATTCAGTTGCAAGGGTAG---AGATATCAGTTTC 1055

Qy      1220 CTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGACCTGAAGCAGCCGGTGCTG 1279
Db      1056 GTGGAATACTATCTCACTAAATATAATATACGCATTCGCGACCACAATCAGCCGCTGCTG 1115

Qy      1280 GTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCTGCCTGGCCCAGCTATGNTC 1339
Db      1116 ATTTCCAAA---AATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTACTA 1172

Qy      1340 ATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAATGCGCAATGATTTCAATGTG 1399
Db      1173 ATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCAGCTT 1232

Qy      1400 ATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCAGCGGCAGCGGGAGGTGGGC 1459
Db      1233 ATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATCGATTGCGC 1292

Qy      1460 CGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAGAGAGCTTCGAGACTGGGGC 1519
Db      1293 GCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGACTGGAAC 1352

Qy      1520 CTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAATCTTACAATCTGAGAAGATC 1579
Db      1353 ATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCAGAACATC 1412

Qy      1580 ----CACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGCAGACTGGTCCAAAGA 1635
Db      1413 GTGTTTCATAATGGAAAGGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTCAGA 1472

Qy      1636 AACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGGCTGCTGATCTATAC 1695
Db      1473 GACCAAAGGATGCTTACCACTCCGAGCGATGGC-CTCGATCGTTGGGCTGTCATCGCGCC 1531

Qy      1696 CCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTTCAAAGTGACTCCAGC 1755
Db      1532 GCAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAGTGG 1591

Qy      1756 CATGGGCATCCAGATGAAAAAGGCAATCA---TGATCGAGGTGGATGACAGAACAGAAGC 1812
Db      1592 AATGGGTCTTAGAATTCGAAGCCCCCAGGAATTCATAATTTATGATGATCGCACTGGAAC 1651

Qy      1813 TTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGTCTCTT 1872
Db      1652 TTATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGCCTCGT 1711

Qy      1873 GTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTGTACAGACTGCCC 1932
Db      1712 ACCCAATGATAACGCCGAAAGATACTCATCAATCAAAAAGAGAGGATACGTTGACAGGGC 1771

Qy      1933 CACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATGGCCAT 1992
Db      1772 GGTGCCAACTCA---AGTTGTGACCCTTAAAACGACCAAGAACCGTAGCCTTATGAGCAT 1828

Qy      1993 TGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGGGTGGA 2052
Db      1829 TGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGATCGA 1888

Qy      2053 CATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACACCACAGCTGG 2112

Db      1889 ACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGATCG 1948

Qy      2113 GCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTC 2172
Db      1949 GAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTT 2008

Qy      2173 CCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGCTTGCA 2232
Db      2009 CAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATGAT 2068

Qy      2233 AGCT---GCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATCGT 2289
Db      2069 AGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATT 2128

Qy      2290 GTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCACA 2349
Db      2129 TTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAAGGA 2188

Qy      2350 GTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATCGT 2409
Db      2189 CATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATTAGC 2248

Qy      2410 GGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAACCC 2469
Db      2249 TTATATTGTGGTAACCAGATCCATGAA---CACGCGCTTCTTCCTCAACGGACAAAATCC 2305

Qy      2470 TCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTCAT 2529
Db      2306 TCCGCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTATCT 2365

Qy      2530 CGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTATGA 2589
Db      2366 GGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAG 2425

Qy      2590 CAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTGTA 2649
Db      2426 CAGCATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTA 2485

Qy      2650 CTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTGGC 2709
Db      2486 CTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGC 2545

Qy      2710 CTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTCTA 2769
Db      2546 TACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTA 2605

Qy      2770 CTACCTCTAA 2779
Db      2606 TTATCTATAA 2615
```

```
RESULT 8
ABL10571
ID   ABL10571 standard; cDNA; 2838 BP.
XX
AC   ABL10571;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 26195.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
```

PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB66468.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 26195; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2838 BP; 783 A; 659 C; 682 G; 714 T; 0 other;

Query Match 7.0%; Score 286; DB 23; Length 2838;
Best Local Similarity 47.6%; Pred. No. 9.3e-35;
Matches 1097; Conservative 0; Mismatches 1168; Indels 42; Gaps 7;

```
Qy   503 AAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCACTTCCGGCTGACCTCG 562
Db   431 AAGAAAGGAGTGGTTGGCACTCATATTACCGTGCAGGCAAACTATTTTAAGGTATTAAAG 490

Qy   563 CGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCCCTGATGGAGGCCCGA 622
Db   491 CGTCCAAACTGGACCATCTACCAGTACCGCGTCGATTTTACGCCTGATGTGGAGGCTACA 550

Qy   623 AGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGAAGGTGTCATGCTTTC 682
Db   551 CGACTGCGACGGTCTTTCTTGTATGAACATAAAGGGATCCTGGGCGGCTACATCTTTGAC 610

Qy   683 GATGGGACAATATT----------GTTTTTACCTAAGAGACTACAGCACAAGGTCACA 730
Db   611 GGAACCAATATGTTTTGCATCAATCAGTTCAAAGCTGTTCAAGATAGCCCCTATGTTTTG 670

Qy   731 GAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCACCCTGACCAAC 790
Db   671 GAACTGGTTACGAAGAGTCGTGCTGGCGAAAACATTGAAATTAAGATCAAGGCTGTTGGA 730

Qy   791 GAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTTCAGGAGGCTC 850
Db   731 TCTGTGCAATCTACGGATGCCGAGCAGTTTCAGGTCCTTAATCTCATACTGCGCAGGGCC 790

Qy   851 TTGAAAATCATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAGTGACCCGATT 910
Db   791 ATGGAGGGCTTAGACTTGAAGCTGGTCTCGCGCTACTACTACGATCCTCAAGCTAAGATT 850

Qy   911 GATATTCCAAACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCATCCTTCAGTAT 970
Db   851 AATTTGGAGAATTTCCGCATGCAATTATGGCCTGGCTATCAGACTTCGATTCGCCAGCAC 910

Qy   971 GAGAACAACATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAGCGAGACTGTC 1030
Db   911 GAAAATGATATATTACTTTGTTCCGAGATATGCCACAAGGTTATGCGAACTGAGACCTTG 970

Qy  1031 CTAGACTTCATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCAGGAGCAAGTG 1090
Db   971 TACAATATTTTGTCCGATGCTATTCGCGACAGTG------ACGATTATCAAAGTACATTT 1024

Qy  1091 TCGAAGGAGCTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGACCTACCGGGTG 1150
Db  1025 AAGCGTGCAGTAATGGGTATGGTAATACTAACGGATTATAATAACAAAACCTATCGCATT 1084

Qy  1151 GATGACATTGACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGATGGCTCGGAG 1210
Db  1085 GACGATGTCGACTTTCAATCGACGCCATTGTGCAAATTTAAAA---CCAATGACGGTGAG 1141

Qy  1211 GTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGACCTGAAGCAG 1270
Db  1142 ATTTCGTACGTGGATTACTATAAGAAGCGATACAACATAATCATTCGCGACTTGAAGCAA 1201

Qy  1271 CCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCTGCCTGGCCCA 1330
Db  1202 CCTCTGGTCATGTCTCGTCCGACAGACAAAAATATTCGTGGTGGAAATGACCAGGCCATA 1261

Qy  1331 GCTATGNTCATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAATGCGCAATGAT 1390
Db  1262 ---ATGATCATTCCCGAGCTGGCACGGGCTACGGGAATGACGGACGCTATGCGCGCTGAC 1318

Qy  1391 TTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCAGCGGCAGCGG 1450
Db  1319 TTTAGGACGTTGAGGGCCATGAGTGAACATACCAGGCTGAATCCAGATCGTCGCATCGAA 1378

Qy  1451 GAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAGAGAGCTTCGA 1510
Db  1379 CGCCTTCGCATGTTCAACAAGCGCTTAAAGTCGTGTAAACAGAGTGTAGAGACCCTTAAG 1438

Qy  1511 GACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAATCTTACAATCT 1570
Db  1439 TCCTGGAACATCGAGCTGGACTCCGCCTTGGTGGAGATTCCAGCACGCGTGTTGCCACCG 1498

Qy  1571 GAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGCAGACTGGTCC 1630
Db  1499 GAAAAAATATTATTCGGCAACCAAAAAATATTCGTATGCGACGCTCGCGCCGATTGGACC 1558

Qy  1631 AAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGGCTGCTGATC 1690
Db  1559 AATGAGTTTCGCACCTGTTCGATGTTTAAAAACGTGCATATCAATAGGTGGTATGTGATC 1618

Qy  1691 TATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTTCAAAGTGACT 1750
Db  1619 ACACCGAGTCGAAATCTGCGCGAGACCCAGGAATTTGTGCAGATGTGCATCCGCACAGCC 1678

Qy  1751 CCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGT---GGATGACAGAACA 1807
Db  1679 AGCAGCATGAAGATGAACATCTGCAATCCAATATATGAGGAAATTCCGGATGACCGCAAC 1738

Qy  1808 GAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGT 1867
Db  1739 GGCACTTACTCCCAAGCGATCGACAACGCCGCAGCTAACGATCCACAGATCGTGATGGTT 1798

Qy  1868 CTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTGTACAGAC 1927
Db  1799 GTCATGAGATCTCCAAATGAAGAGAAATATAGCTGCATTAAAAAACGCACGTGCGTGGAC 1858

Qy  1928 TGCCCCACCCCAAGTCAGTGTGTG----------GTGGCCCGGACCCTGGGCAAGCAG 1975
Db  1859 AGACCGGTGCCATCGCAGGTGGTGACACTAAAAGTTATCGCGCCTCGACAGCAAAAACCA 1918

Qy  1976 CAAACAGTCATGGCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGC 2035
Db  1919 ACTGGGCTGATGTCGATCGCCACAAAGGTGGTTATTCAAATGAACGCCAAATTGATGGGA 1978

Qy  2036 GAGCTCTGGCGGGTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTAC 2095
Db  1979 GCTCCCTGGCAGGTAGTGATCCCCCTCCACGGTCTGATGACTGTTGGTTTCGATGTCTGC 2038

Qy  2096 CATGACACCACAGCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCAT---CAATGAA 2152
Db  2039 CATTCACCGAAGAATAAGAACAAGTCATATGGGGCTTTCGTTGCAACCATGGACCAGAAG 2098
```

```
Qy       2153 GGGATGACCCGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGAT 2212
Db       2099 GAGTCTTTCCGCTACTTCTCCACCGTAAACGAACACATTAAGGGCCAAGAGTTGTCCGAG 2158

Qy       2213 GGTCTCAAGGTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATG 2272
Db       2159 CAGATGTCGGTGAACATGGCGTGCGCCCTGAGGTCGTATCAGGAGCAACACCGTTCTTTG 2218

Qy       2273 CCCAGCCGTGTCATCGTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTC 2332
Db       2219 CCAGAGCGCATTCTCTTCTTTCGCGACGGTGTTGGCGATGGTCAGCTCTACCAGGTGGTA 2278

Qy       2333 AATTATGAGGTCCCACAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCA 2392
Db       2279 AACAGCGAGGTGAACACCCTAAAGGACAGGCTCGACGAAATTTACAAATCAGCTGGCAAA 2338

Qy       2393 AGACTGACTGTAATCGTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGG 2452
Db       2339 CAGGAGGGCTGTCGCATGACATTTATAATTGTATCCAAGCGCATTAATTCTCGCTACTTT 2398

Qy       2453 GGAAGACTTCAGAACCCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAG 2512
Db       2399 ACTGGGCATCGCAACCCAGTTCCGGGCACTGTAGTCGATGACGTTATTACCTTGCCAGAG 2458

Qy       2513 TGGTATGACTTTTTCATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACAC 2572
Db       2459 CGCTACGACTTCTTCCTAGTGTCCCAGGCTGTTCGGATAGGAACTGTGTCGCCTACCAGC 2518

Qy       2573 TACAATGTCATCTATGACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATAC 2632
Db       2519 TACAATGTTATTTCTGACAACATGGGACTAAACGCCGATAAGCTGCAGATGCTCTCCTAT 2578

Qy       2633 AAGNTCTGCCACGTGTACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAG 2692
Db       2579 AAGATGACCCATATGTACTACAATTACAGCGGAACCATACGAGTCCCCGCTGTCTGCCAT 2638

Qy       2693 TATGCACACAAGCTGGCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCC 2752
Db       2639 TACGCCCACAAATTGGCTTTCCTCGTAGCCGAATCCATTAATCGTGCGCCTTCAGCAGGA 2698

Qy       2753 CTGTCCAACCGCCTCTACTACCTCTAA 2779
Db       2699 CTGCAGAATCAATTGTACTTTTTGTAA 2725
```

```
RESULT 9
AAH14032
ID   AAH14032 standard; cDNA; 1733 BP.
XX
AC   AAH14032;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:11143.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 11143; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 1733 BP; 435 A; 408 C; 451 G; 439 T; 0 other;

  Query Match          6.4%;  Score 259.6;  DB 22;  Length 1733;
  Best Local Similarity  53.2%;  Pred. No. 9.3e-31;
  Matches  622;  Conservative    0;  Mismatches  535;  Indels   12;  Gaps    3;

Qy       1621 AGACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTG 1680
Db          3 AAACTGGGTTAAGGAAGTAACCAGAGACCCTTCCATCTTGACTATCCCCATGCATTTCTG 62

Qy       1681 GCTGCTGATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTT 1740
Db         63 GGCACTTTTTTACCCAAAGAGAGCAATGGACCAGGCTCGAGAACTGGTCAACATGTTGGA 122

Qy       1741 CAAAGTGACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGT---GGA 1797
Db        123 GAAGATAGCCGGCCCCATTGGCATGCGTATGAGCCCACCGGCCTGGGTTGAACTAAAGGA 182

Qy       1798 TGACAGAACAGAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAG------ACAC 1851
Db        183 TGACCGAATAGAGACTTATGTCAGAACCATTCAATCCACGTTAGGAGCTGAGGGGAAGAT 242

Qy       1852 TCAGATAGTTGTCTGTCTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAA 1911
Db        243 ACAGATGGTTGTTTGCATCATCATGGGCCCACGTGATGATCTCTATGGGGCCATCAATAA 302

Qy       1912 GTACTTGTGTACAGACTGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAA 1971
Db        303 GCTGTGCTGTGTGCAGTCCCCAGTGCCCTCCCAGGTTGTCAATGTTCGAACCATTGGTCA 362

Qy       1972 GCAGCAAACAGTCATGGCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGG 2031
Db        363 GCCCACCAGGCTTCGGAGTGTGGCCCAGAAGATTTTACTTCAGATTAACTGTAAATTGGG 422

Qy       2032 AGGCGAGCTCTGGCGGGTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTG 2091
Db        423 TGGTGAGCTCTGGGGAGTGGATATTCCTCTGAAACAGTTAATGGTGATCGGGATGGATGT 482
```

```
Qy       2092 TTACCATGACACCACAGCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGA 2151
Db        483 TTACCATGACCCCAGTAGAGGCATGCGCTCCGTGGTTGGCTTCGTGGCAAGCATCAATCT 542

Qy       2152 AGGGATGACCCGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGA 2211
Db        543 CACCCTCACAAAATGGTATTCCCGGGTGGTGTTCCAGATGCCGCATCAGGAGATTGTGGA 602

Qy       2212 TGGTCTCAAGGTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACAT 2271
Db        603 CAGCCTGAAGCTATGCCTCGTGGGCTCCTTAAAAAAGTTTTATGAGGTGAACCACTGTCT 662

Qy       2272 GCCCAGCCGTGTCATCGTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGT 2331
Db        663 ACCAGAGAAGATTGTGGTGTACCGTGATGGAGTGTCTGATGGCCAACTGAAGACAGTTGC 722

Qy       2332 CAATTATGAGGTCCCACAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCC 2391
Db        723 CAACTATGAGATTCCTCAACTACAGAAGTGTTTTGAAGCTTTTGAGAA---TTATCAGCC 779

Qy       2392 AAGACTGACTGTAATCGTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGG 2451
Db        780 CAAGATGGTGGTGTTTGTAGTTCAGAAGAAAATCAGTACTAATCTATATCTGGCTGCTCC 839

Qy       2452 GGGAAGACTTCAGAACCCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGA 2511
Db        840 TCAGAACTTTGTAACTCCCACTCCTGGAACTGTGGTAGATCATACAATAACAAGCTGTGA 899

Qy       2512 GTGGTATGACTTTTTCATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACA 2571
Db        900 GTGGGTGGATTTCTATCTTCTTGCCCATCATGTACGGCAGGGCTGTGGCATTCCTACGCA 959

Qy       2572 CTACAATGTCATCTATGACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATA 2631
Db        960 TTATGTCTGTGTTCTCAACACCGCAAACCTGAGCCCTGATCATATGCAGAGGCTGACTTT 1019

Qy       2632 CAAGNTCTGCCACGTGTACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCA 2691
Db        1020 CAAACTGTGCCACATGTACTGGAATTGGCCTGGCACCATCAGAGTTCCAGCTCCTTGCAA 1079

Qy       2692 GTATGCACACAAGCTGGCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTC 2751
Db        1080 GTATGCCCACAAGCTAGCTTTCCTGTCAGGACACATCTTGCATCATGAACCAGCCATCCA 1139

Qy       2752 CCTGTCCAACCGCCTCTACTACCTCTAAC 2780
Db        1140 GCTGTGCGAGAACCTGTTCTTCCTGTGAC 1168
```

RESULT 10
AAZ13404
ID AAZ13404 standard; cDNA; 300 BP.
XX
AC AAZ13404;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:873.
XX
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9938972-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01619.
XX
PR 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI; 1999-494092/41.
XX
PT Novel human genes and their expression products which are
PT differentially expressed in different cell types
XX
PS Claim 1; Page 862-863; 2479pp; English.
XX
CC The present invention describes a library of human polynucleotides
CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
SQ Sequence 300 BP; 75 A; 70 C; 91 G; 64 T; 0 other;

Query Match 5.7%; Score 231.2; DB 20; Length 300;
Best Local Similarity 88.4%; Pred. No. 1.8e-26;
Matches 251; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
Qy       1936 CCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATGGCCATTGC 1995
Db          1 CCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGC 60

Qy       1996 CACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGGGTGGACAT 2055
Db         61 TACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACAT 120

Qy       2056 GGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACACCACAGCTGGGCG 2115
Db        121 CCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCG 180

Qy       2116 GAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCCCG 2175
Db        181 GAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACG 240

Qy       2176 CTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCA 2219
Db        241 CTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCA 284
```

RESULT 11
AAS70144
ID AAS70144 standard; cDNA; 1026 BP.
XX

```
AC   AAS70144;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #5948.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG05957.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 5948; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1026 BP; 282 A; 249 C; 259 G; 236 T; 0 other;

 Query Match             4.6%;  Score 185.4;  DB 23;  Length 1026;
 Best Local Similarity   59.9%;  Pred. No. 1.7e-19;
 Matches  328;  Conservative    0;  Mismatches  217;  Indels    3;  Gaps    1;

Qy        936 TCTGGCCCGGCTTCACCACCTCCATCCTTCAGTATGAGAACAACATCATGCTCTGCACAG 995
              | ||||| || ||  |||  ||  |      | | ||| |  ||  || || |  |  | |
Db        185 TTTGGCCTGGGTTTGCCATTTCTGTGTCATATTTTGAAAGGAAGCTCCTGTTTAGTGCTG 244

Qy        996 ACGTCAGCCACAAGGTGCTCCGCAGCGAGACTGTCCTAGACTTCATGTTCAATCTATACC 1055
              | || ||  |||| || ||||| |  ||||| || || || ||||||     ||| |  |
Db        245 ATGTGAGTTACAAAGTCCTCCGGAATGAGACGGTTCTGGAATTCATGACTGCTCTCTGTC 304

Qy       1056 AGCAGACAGAGGAGCACAAGTTCCAGGAGCAAGTGTCGAAGGAGCTCATAGGCCTCATCG 1115
              |    || |    |  |   |||    ||        |||| |||| ||||| ||||| |
Db        305 AAAGAACTGGCTTGTCCTGTTTCACCCAGACGTGTGAGAAGCAGCTAATAGGGCTCATTG 364

Qy       1116 TTCTCACCAAGTACAATAACAAGACCTACCGGGTGGATGACATTGACTGGGACCAGAATC 1175
              | || || |  |||||||||  ||||||    | ||||||||||||||     ||| |
Db        365 TCCTTACAAGATACAATAACAGAACCTACTCCATTGATGACATTGACTGGTCAGTGAAGC 424

Qy       1176 CAAAGAGCACCTTCAAGAAGGCGGATGGCTCGGAGGTCAGCTTCCTGGAGTACTACAGGA 1235
              | |    ||||||  |||||  ||||||| | ||| ||| ||   |||| |||||||
Db        425 CCACACACACCTTTCAGAAGCGGGATGGCACCGAGATCACCTATGTGGATTACTACAAGC 484

Qy       1236 AGCAATACAACCAGGAGATCACGGACCTGAAGCAGCCGGTGCTGGTGAGCCAACCCAAGC 1295
              |||| ||  |         |  |||||||||| |||||  |||| || || |     |||
Db        485 AGCAGTATGATATTACTGTATCGGACCTGAATCAGCCCATGCTTGTTAGTCTGTTAAAGA 544

Qy       1296 GGAGGAGAGGCCCCGGCGGCACCCTGCCTGGCCCAGCTATGNTCATCCCTGAACTCTGCT 1355
              || ||||      | |  ||    | ||   |  ||      | || ||||| |||||||
Db        545 AGAAGAGAAAT---GACAACAGTGAGGCTCAGCTCGCCCACCTGATACCTGAGCTCTGCT 601

Qy       1356 ATCTCACAGGCCTGACTGATAAAATGCGCAATGATTTCAATGTGATGAAGGACCTGGCAG 1415
              ||| ||||| ||||||||  |          |||||| |  |||||||||   |||| |
Db        602 TTCTAACAGGGCTGACTGACCAGGCAACATCTGATTTCCAGCTGATGAAGGCTGTGGCTG 661

Qy       1416 TGCACACGCGGCTGACCCCTGAGCAGCGGCAGCGGGAGGTGGGCCGCCTCATCGACTACA 1475
              | || || || |  |||      |||||||  |    ||| | | ||  | ||| |||
Db        662 AAAAGACACGTCTCAGTCCTTCAGGCCGGCAGCAGCGCCTGGCCAGGCTTGTGGACAACA 721

Qy       1476 TCCACAAG 1483
              |||| | |
Db        722 TCCAGAGG 729


RESULT 12
ABL10570
ID   ABL10570 standard; cDNA; 5671 BP.
XX
AC   ABL10570;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 26192.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB66467.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 26192; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
```

```
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 5671 BP; 1723 A; 1127 C; 1131 G; 1690 T; 0 other;

  Query Match             4.3%;  Score 175;  DB 23;  Length 5671;
  Best Local Similarity   52.0%;  Pred. No. 6.8e-18;
  Matches  416;  Conservative    0;  Mismatches  381;  Indels    3;  Gaps    1;

Qy       1983 TCATGGCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCT 2042
             | ||| | || ||||| ||| | |   | || ||||||  |||  ||   || |  | ||
Db       3759 TGATGTCGATCGCCACAAAGGTGGTTATTCAAATGAACGCCAAATTGATGGGAGCTCCCT 3818

Qy       2043 GGCGGGTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACA 2102
             ||| ||| |  ||  ||||  |       ||||  || ||  ||||    | ||||
Db       3819 GGCAGGTAGTGATCCCCCTCCACGGTCTGATGACTGTTGGTTTCGATGTCTGCCATTCAC 3878

Qy       2103 CCACAGCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGAC-- 2160
             | |    |  |   | |||    |  |  ||||| || | |||  |  |   |  | |
Db       3879 CGAAGAATAAGAACAAGTCATATGGGGCTTTCGTTGCAACCATGGACCAGAAGGAGTCTT 3938

Qy       2161 -CCGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCA 2219
              |||||  ||||||  |     |  ||       || || ||| ||  ||     |
Db       3939 TCCGCTACTTCTCCACCGTAAACGAACACATTAAGGGCCAAGAGTTGTCCGAGCAGATGT 3998

Qy       2220 AGGTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCC 2279
              ||||  | ||      || |||||| | |     |       |    |   ||||    |
Db       3999 CGGTGAACATGGCGTGCGCCCTGAGGTCGTATCAGGAGCAACACCGTTCTTTGCCAGAGC 4058

Qy       2280 GTGTCATCGTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATG 2339
             |  |  || | |  || |||||||| || || || |||||  |     |||| ||    |
Db       4059 GCATTCTCTTCTTTCGCGACGGTGTTGGCGATGGTCAGCTCTACCAGGTGGTAAACAGCG 4118

Qy       2340 AGGTCCCACAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGA 2399
              ||||       |    ||  | ||| |   | |    | |       |  |     |
Db       4119 AGGTGAACACCCTAAAGGACAGGCTCGACGAAATTTACAAATCAGCTGGCAAACAGGAGG 4178

Qy       2400 CTGTAATCGTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGAC 2459
                | ||      |    |||         |  ||  | |||  |         | |
Db       4179 GCTGTCGCATGACATTTATAATTGTATCCAAGCGCATTAATTCTCGCTACTTTACTGGGC 4238

Qy       2460 TTCAGAACCCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATG 2519
              ||  |||||  |||| || || ||  ||||||  |   | |||   ||||||  | || |
Db       4239 ATCGCAACCCAGTTCCGGGCACTGTAGTCGATGACGTTATTACCTTGCCAGAGCGCTACG 4298

Qy       2520 ACTTTTTCATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATG 2579
              |||| ||| | |||  |||||| ||  | |  || | |||||| || ||   ||||||||
Db       4299 ACTTCTTCCTAGTGTCCCAGGCTGTTCGGATAGGAACTGTGTCGCCTACCAGCTACAATG 4358

Qy       2580 TCATCTATGACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCT 2639
             | || | ||||| ||  ||  || ||  ||||  |  | |||  |||  | || ||| |
Db       4359 TTATTTCTGACAACATGGGACTAAACGCCGATAAGCTGCAGATGCTCTCCTATAAGATGA 4418

Qy       2640 GCCACGTGTACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCAC 2699
              |||  |||||||  ||||      |||  ||| ||||||||  ||    |||| || || |
Db       4419 CCCATATGTACTACAATTACAGCGGAACCATACGAGTCCCCGCTGTCTGCCATTACGCCC 4478

Qy       2700 ACAAGCTGGCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCA 2759
              ||||  |||| ||||||||  | | |   |||  |  | | |||          |||   |
Db       4479 ACAAATTGGCTTTCCTCGTAGCCGAATCCATTAATCGTGCGCCTTCAGCAGGACTGCAGA 4538

Qy       2760 ACCGCCTCTACTACCTCTAA 2779
              | |   | ||||   | |||
Db       4539 ATCAATTGTACTTTTTGTAA 4558


RESULT 13
ABL10534
ID   ABL10534 standard; cDNA; 7667 BP.
XX
AC   ABL10534;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 26084.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB66431.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 26084; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 7667 BP; 2236 A; 1513 C; 1554 G; 2364 T; 0 other;

  Query Match             3.9%;  Score 159.4;  DB 23;  Length 7667;
  Best Local Similarity   51.1%;  Pred. No. 1.6e-15;
  Matches  426;  Conservative    0;  Mismatches  402;  Indels    6;  Gaps    2;

Qy       1949 GTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATGGCCATTGCCACCAAGATCGCC 2008
              || ||| |||  |    |  |||| | |  |   | |||  ||||||||||| || ||
Db       4928 GTTGTGACCCTTAAAACGACCAAGAACCGTAGCCTTATGAGCATTGCCACCAAAATAGCA 4987

Qy       2009 CTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGGGTGGACATGGCCCTGAAACTG 2068
              | ||  |||| |||||| |||||       | ||||  | | ||  |  || ||
Db       4988 ATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGATCGAACTACCCTTGTCCGGA 5047

Qy       2069 GCAATGATCGTGGGCATCGACTGTTACCATGACACCACAGCTGGGCGGAGGTCCATCGCA 2128
                ||||   | ||| | |||  |    |   |||   || | ||  |||| ||  || |
Db       5048 CTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGATCGGAAGAGGGCCTACGGA 5107

Qy       2129 GGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCCCGCTGCGTCTTTCAG 2188
              | |||  | |||   ||  ||  |  |  | |  | || || | ||   |||    ||
Db       5108 GCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTTCAGCACAGTCACGGAG 5167

Qy       2189 GACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGCTTGCAAGC---TGCTCTGAGG 2245
                | |||     ||  || | | |     |     |    ||  |||    || ||| |
Db       5168 TGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATGATAGCAAAGGCCCTGCGC 5227

Qy       2246 GCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATCGTGTACCGAGACGGTGTG 2305
                 |            ||    | ||||    ||  || |  | || ||||||||||||
```

```
Db       5228 CAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATTTTATCGAGACGGTGTG 5287

Qy       2306 GGGGACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCACAGTTCCTAGATTGCCTC 2365

Db       5288 AGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAAGGACATCATTGAGAAGTTG 5347

Qy       2366 AAGTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATCGTGGTGAAGAAGCGTGTC 2425

Db       5348 AAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATTAGCTTATATTGTGGTAACC 5407

Qy       2426 AATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAACCCTCTTCCAGGGACAGTC 2485

Db       5408 AGATCCATGAACA---CGCGCTTCTTCCTCAACGGACAAAATCCTCCGCCTGGTACTATA 5464

Qy       2486 ATCGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTCATCGTGAGCCAGGCAGTG 2545

Db       5465 GTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTATCTGGTCTCGCAACAAGTT 5524

Qy       2546 AGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTATGACAGCAGTGGCCTGAAG 2605

Db       5525 CGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAGCAGCATGGGTCTCTCA 5584

Qy       2606 CCCGACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTGTACTATAATTGGCCTGGA 2665

Db       5585 CCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTACTACAATTGGTCGGGC 5644

Qy       2666 GTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTGGCCTTCCTCGTGGGCCAG 2725

Db       5645 ACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGCTACCCTCGTGGGTACG 5704

Qy       2726 AGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTCTACTACCTCTAA 2779

Db       5705 AACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTATTATCTATAA 5758
```

RESULT 14
ABN62610
ID ABN62610 standard; cDNA; 548 BP.
XX
AC ABN62610;
XX
DT 28-JUN-2002 (first entry)
XX
DE Human cancer related polynucleotide SEQ ID NO 2577.
XX
KW Human; cytostatic; gene expression; gene mapping; tissue profiling;
KW gene therapy; cancer; tumour; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200214500-A2.
XX
PD 21-FEB-2002.
XX
PF 16-AUG-2001; 2001WO-US25840.
XX
PR 16-AUG-2000; 2000US-226326P.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;
PI Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;
XX
DR WPI; 2002-241905/29.
XX
PT New nucleic acid for producing a polypeptide, detecting differentially
PT expressed genes correlated with a cancerous state of a mammalian cell,
PT and inhibiting tumor growth -
XX
PS Claim 1; SEQ ID NO 2577; 883pp + Sequence Listing; English.
XX
CC The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
CC with cytostatic activity. The polynucleotide is used to produce a
CC polypeptide, to detect differentially expressed genes correlated with a
CC cancerous state of a mammalian cell and to inhibit tumour growth. The
CC polynucleotide is used as a probe in mapping and tissue profiling. The
CC encoded polypeptide and antibodies to the polypeptide can also be used
CC for therapeutic and diagnostic purposes. The polynucleotide is useful for
CC gene therapy.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 548 BP; 151 A; 116 C; 146 G; 134 T; 1 other;

 Query Match 3.9%; Score 157.6; DB 24; Length 548;
 Best Local Similarity 55.9%; Pred. No. 2.7e-15;
 Matches 298; Conservative 0; Mismatches 235; Indents 0; Gaps 0;

```
Qy       1970 AAGCAGCAAACAGTCATGGCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATG 2029

Db          5 AAACAGGGCATGATGATGAGTATCGCCACCAAGATCGCTATGCAGATGACTTGCAAGCTC 64

Qy       2030 GGAGGCGAGCTCTGGCGGGTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGAC 2089

Db         65 GGAGGCGAGCTGTGGGCTGTGGAAATACCTTTAAAGTCCCTGATGGTGGTCGGTATTGAT 124

Qy       2090 TGTTACCATGACACCACAGCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAAT 2149

Db        125 GTCTGTAAAGATGCACTCAGCAAGGACGCGATGGTTGTTGGATGCGTGGCCAGTGTTAAC 184

Qy       2150 GAAGGGATGACCCGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTG 2209

Db        185 CCCAGAATCACCAGGTGGTTTTCCCGCTGTATCCTTCAGAGAACAATGACTGATGTTGCA 244

Qy       2210 GATGGTCTCAAGGTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATAC 2269

Db        245 GATTGCTTGAAAGTTTTCATGACTGGAGCACTCAACAAATGGTACAAGTACAATCATGAT 304

Qy       2270 ATGCCCAGCCGTGTCATCGTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTG 2329

Db        305 TTGCCAGCACGGATAATTGTGTACCGTGCTGGTGTAGGGGATGGTCAGCTGAAAACACTT 364

Qy       2330 GTCAATTATGAGGTCCCACAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAAC 2389

Db        365 ATTGAATATGAAGTCCCACAGCTGCTGAGCAGTGTGGCAGAATCCAGCTCANATACCAGC 424

Qy       2390 CCAAGACTGACTGTAATCGTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCT 2449

Db        425 TCAAGACTGTCGGTGATTGTGGTCAGGAAGAAGTGCATGCCACGATTCTTTACCGAAATG 484

Qy       2450 GGGGGAAGACTTCAGAACCCTCTTCCAGGGACAGTCATCGATGTGGAAGTCAC 2502

Db        485 AACCGCACTGTACAGAACCCCCCACTTGGCACTGTTGTGGATTCAGAAGCAAC 537
```

RESULT 15
AAI88514
ID AAI88514 standard; cDNA; 389 BP.
XX
AC AAI88514;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 8574.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX

PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR P-PSDB; AAO08583.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 1; SEQ ID NO 8574; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 389 BP; 259 A; 15 C; 49 G; 66 T; 0 other;

Query Match 3.1%; Score 127.4; DB 22; Length 389;
Best Local Similarity 87.0%; Pred. No. 1e-10;
Matches 140; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```
Qy       3904 TGTGTGTGTGCCCCTGGCACAGCTCAATAAAGTTTCTGTCTCACTTTAAAAAAAAAAAAA 3963
             ||| ||  | ||| || |     | |||||||||  | | |   |  ||||||||||||
Db         40 TGTTTGCATCCCCTTGCCCTCAATAAATAAAGTTCTTTTTTAGTTCCAAAAAAAAAAAAA 99

Qy       3964 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4023
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        100 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 159

Qy       4024 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4064
             ||||||||||||||||||||||||||||||||||||||||
Db        160 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 200
```

Search completed: July 21, 2003, 19:01:21
Job time : 814.65 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2003, 18:06:22 ; Search time 165.893 Seconds
(without alignments)
7512.871 Million cell updates/sec

Title: US-09-873-737A-3
Perfect score: 4064
Sequence: 1 agaagcctgctacaggggt.....aaaaaaaaaaaaaaaaaaa 4064

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	120.4	3.0	1798	4	US-09-797-906-1	Sequence 1, Appli
2	118.8	2.9	1813	4	US-09-071-224-3	Sequence 3, Appli
3	118.8	2.9	2082	2	US-08-785-310A-2	Sequence 2, Appli
4	118.6	2.9	2674	4	US-09-817-180-1	Sequence 1, Appli
5	118.4	2.9	3275	4	US-09-370-838-151	Sequence 151, App
6	118	2.9	2447	2	US-09-014-969-14	Sequence 14, Appl
7	117.8	2.9	240	1	US-08-628-417-6	Sequence 6, Appli
8	117.6	2.9	144	1	US-08-702-344-26	Sequence 26, Appl
9	117.4	2.9	1733	3	US-09-073-569-1	Sequence 1, Appli
10	115.6	2.8	1117	4	US-09-247-373B-33	Sequence 33, Appl
11	114.2	2.8	2246	4	US-09-363-708-3	Sequence 3, Appli
12	114	2.8	117	1	US-08-702-344-3	Sequence 3, Appli
13	114	2.8	1051	4	US-09-245-041-10	Sequence 10, Appl
14	114	2.8	1447	4	US-09-443-041A-27	Sequence 27, Appl
15	114	2.8	2269	4	US-09-394-645-1	Sequence 1, Appli
16	114	2.8	2269	4	US-09-243-560B-1	Sequence 1, Appli
17	111.6	2.7	2634	4	US-09-463-238-3	Sequence 3, Appli
18	111.2	2.7	6671	1	US-08-280-443-1	Sequence 1, Appli
19	111.2	2.7	6671	1	US-08-457-459-1	Sequence 1, Appli
20	111.2	2.7	6671	1	US-08-555-678-1	Sequence 1, Appli
21	111.2	2.7	6671	5	PCT-US95-02275-1	Sequence 1, Appli
22	110.8	2.7	2186	4	US-09-360-545-66	Sequence 66, Appl
23	109.2	2.7	2184	4	US-08-955-918C-1	Sequence 1, Appli
24	109.2	2.7	2184	4	US-08-697-766A-1	Sequence 1, Appli
25	109	2.7	111	4	US-09-297-535-23	Sequence 23, Appl
26	109	2.7	121	4	US-09-297-535-20	Sequence 20, Appl
27	108.8	2.7	140	1	US-08-628-417-5	Sequence 5, Appli
c 28	108	2.7	9589	1	US-07-925-695-1	Sequence 1, Appli
c 29	108	2.7	9589	1	US-07-925-695-2	Sequence 2, Appli
30	107.8	2.7	578	4	US-09-602-877A-95	Sequence 95, Appl
31	107.8	2.7	1066	1	US-08-157-101A-4	Sequence 4, Appli
32	107.4	2.6	1872	4	US-09-801-052-1	Sequence 1, Appli
33	106.2	2.6	1882	4	US-09-370-253-1	Sequence 1, Appli
34	104.4	2.6	1474	4	US-08-821-994-64	Sequence 64, Appl
35	101.6	2.5	1454	4	US-09-372-422A-19	Sequence 19, Appl
36	98.8	2.4	2323	4	US-09-149-476-24	Sequence 24, Appl
c 37	98.6	2.4	12980	3	US-08-811-566-5	Sequence 5, Appli
c 38	98.6	2.4	12980	4	US-09-034-756-5	Sequence 5, Appli
39	98	2.4	98	1	US-08-088-658-42	Sequence 42, Appl
40	98	2.4	98	2	US-08-471-907A-42	Sequence 42, Appl
c 41	98	2.4	356	2	US-08-520-678A-22	Sequence 22, Appl
c 42	98	2.4	356	4	US-08-897-126-22	Sequence 22, Appl
c 43	98	2.4	9646	3	US-08-811-566-1	Sequence 1, Appli
c 44	98	2.4	9646	4	US-09-034-756-1	Sequence 1, Appli
45	97	2.4	790	4	US-09-363-970-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-797-906-1
; Sequence 1, Application US/09797906
; Patent No. 6329188
; GENERAL INFORMATION:
; APPLICANT: Zianghe YAN, Karen A. KETCHUM, Valentina DIFRANCESCO, Ellen M. BEASLEY
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001151CIP
; CURRENT APPLICATION NUMBER: US/09/797,906
; CURRENT FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1798
; TYPE: DNA
; ORGANISM: Human
US-09-797-906-1

Query Match 3.0%; Score 120.4; DB 4; Length 1798;
Best Local Similarity 89.0%; Pred. No. 3e-16;
Matches 130; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```
Qy       3919 GGCACAGCTCAATAAAGTTTCTGTCTCACTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3978
              ||||  | | |  ||  | ||  |  || |  ||||||||||||||||||||||||||||
Db       1613 GGCATTGATTACAAACATGTCCTTGGCAGTGGAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1672

Qy       3979 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4038
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1673 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1732

Qy       4039 AAAAAAAAAAAAAAAAAAAAAAAA 4064
              ||||||||||||||||||||||||
Db       1733 AAAAAAAAAAAAAAAAAAAAAAAA 1758
```

RESULT 2
US-09-071-224-3
; Sequence 3, Application US/09071224
; Patent No. 6271343
; GENERAL INFORMATION:
; APPLICANT: Lok, Si
; APPLICANT: Presnell, Scott R.
; APPLICANT: Jelmberg, Anna C.
; APPLICANT: Gilbert, Teresa
; APPLICANT: Foster, Donald C.
; APPLICANT: Adams, Robyn L.
; APPLICANT: Lehner, Joyce M.
; TITLE OF INVENTION: MAMMALIAN ZCYTOR5

```
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Zymogenetics
;      STREET:  1201 Eastlake Ave East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,224
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lunn, Paul G
;      REGISTRATION NUMBER:  32,743
;      REFERENCE/DOCKET NUMBER:  96-22
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6627
;      TELEFAX:  206-442-6678
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1813 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  88...1362
;      OTHER INFORMATION:
US-09-071-224-3

  Query Match             2.9%;  Score 118.8;  DB 4;  Length 1813;
  Best Local Similarity   91.3%;  Pred. No. 6.5e-16;
  Matches  126;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       3927 TCAATAAAGTTTCTGTCTCACTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3986
              | ||||||    || | |   | ||||||||||||||||||||||||||||||||||||
Db       1664 TAAATAAATGAGCTATTTAGGTGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1723

Qy       3987 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1724 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1783

Qy       4047 AAAAAAAAAAAAAAAAA 4064
              |||||||||||||||||
Db       1784 AAAAAAAAAAAAAAAAA 1801


RESULT 3
US-08-785-310A-2
; Sequence 2, Application US/08785310A
; Patent No. 5840532
;  GENERAL INFORMATION:
;    APPLICANT:  McKnight, Steven L.
;    APPLICANT:  Russell, David W.
;    TITLE OF INVENTION:  Neuronal PAS Domain Protein
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 BUSH STREET, SUITE 3200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/785,310A
;      FILING DATE:  21-JAN-1997
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UTSD:1226
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 343-4341
;      TELEFAX:  (415) 343-4342
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2082 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-785-310A-2

  Query Match             2.9%;  Score 118.8;  DB 2;  Length 2082;
  Best Local Similarity   94.6%;  Pred. No. 6.8e-16;
  Matches  123;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       3935 GTTTCTGTCTCACTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3994
              | | ||| ||| |   ||||||||||||||||||||||||||||||||||||||||||||
Db       1947 GATGCTGCCTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2006

Qy       3995 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4054
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2007 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2066

Qy       4055 AAAAAAAAAA 4064
              ||||||||||
Db       2067 AAAAAAAAAA 2076


RESULT 4
US-09-817-180-1
; Sequence 1, Application US/09817180
; Patent No. 6340584
; GENERAL INFORMATION:
;  APPLICANT: GAN, Weiniu et al.
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001183
;  CURRENT APPLICATION NUMBER: US/09/817,180
;  CURRENT FILING DATE:  2001-03-27
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 2674
;   TYPE: DNA
;   ORGANISM: Human
US-09-817-180-1

  Query Match             2.9%;  Score 118.6;  DB 4;  Length 2674;
  Best Local Similarity   87.2%;  Pred. No. 8.3e-16;
  Matches  130;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy       3916 CCTGGCACAGCTCAATAAAGTTTCTGTCTCACTTTAAAAAAAAAAAAAAAAAAAAAAAAA 3975
              || |||| |    |  |||     |||  |   | ||||||||||||||||||||||||
Db       2512 CCGGGCAGAAACAATAAAACCACTTGTGCCCACTGAAAAAAAAAAAAAAAAAAAAAAAAA 2571

Qy       3976 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4035
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2572 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2631

Qy       4036 AAAAAAAAAAAAAAAAAAAAAAAAAAAA 4064
              ||||||||||||||||||||||||||||
Db       2632 AAAAAAAAAAAAAAAAAAAAAAAAAAAA 2660


RESULT 5
US-09-370-838-151
; Sequence 151, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
;  APPLICANT: Reed, Steven G.
;  APPLICANT: Lodes, Michael J.
;  APPLICANT: Mohamath, Roadoh
;  APPLICANT: Secrist, Heather
;  TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
;  TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.475C1
;  CURRENT APPLICATION NUMBER: US/09/370,838
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/285,323
;  EARLIER FILING DATE: 1999-04-02
;  NUMBER OF SEQ ID NOS: 289
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 151
;   LENGTH: 3275
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-370-838-151

  Query Match             2.9%;  Score 118.4;  DB 4;  Length 3275;
  Best Local Similarity   91.9%;  Pred. No. 9.8e-16;
  Matches  125;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       3929 AATAAAGTTTCTGTCTCACTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3988
              || |||||| |   |  |    ||||||||||||||||||||||||||||||||||||||||
Db       3114 AAAAAAGTTACATGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3173

Qy       3989 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4048
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3174 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3233

Qy       4049 AAAAAAAAAAAAAAAA 4064
              ||||||||||||||||
Db       3234 AAAAAAAAAAAAAAAA 3249


RESULT 6
US-09-014-969-14
; Sequence 14, Application US/09014969
; Patent No. 5965397
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John M.
;    APPLICANT:  LaVallie, Edward R.
;    APPLICANT:  Racie, Lisa A.
;    APPLICANT:  Merberg, David
;    APPLICANT:  Treacy, Maurice
;    APPLICANT:  Spaulding, Vikki
;    APPLICANT:  Agostino, Michael J.
;    TITLE OF INVENTION:  SECRETED PROTEINS AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  ENCODING THEM
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  MA
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/014,969
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sprunger, Suzanne A.
;      REGISTRATION NUMBER:  41,323
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8284
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2447 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-014-969-14

  Query Match             2.9%;  Score 118;  DB 2;  Length 2447;
  Best Local Similarity   91.2%;  Pred. No. 1.1e-15;
  Matches  124;  Conservative    1;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       3929 AATAAAGTTTCTGTCTCACTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3988
              ||||||  || ||    |    |||||||||:||||||||||||||||||||||||||||||||
Db       2228 AATAAACCTTTTGGGAGAAAAAAAAAAAAAAARAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2287

Qy       3989 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4048
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2288 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2347

Qy       4049 AAAAAAAAAAAAAAAA 4064
              ||||||||||||||||
Db       2348 AAAAAAAAAAAAAAAA 2363


RESULT 7
US-08-628-417-6
; Sequence 6, Application US/08628417
; Patent No. 5627054
;  GENERAL INFORMATION:
;    APPLICANT:  GILLESPIE, DAVID
;    TITLE OF INVENTION:  COMPETITOR PRIMER ASYMMETRIC
;    TITLE OF INVENTION:  POLYMERASE CHAIN REACTION
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  U.S. ARMY CHEMICAL AND BIOLOGICAL
;      ADDRESSEE:  DEFENSE COMMAND
;      STREET:  OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)
;      CITY:  ABERDEEN PROVING GROUND
;      STATE:  MARYLAND
;      COUNTRY:  USA
;      ZIP:  21010-5423
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/628,417
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BIFFONI, ULYSSES J
;      REGISTRATION NUMBER:  39,908
;      REFERENCE/DOCKET NUMBER:  DAM 398-94
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  410-671-1158
;      TELEFAX:  410-671-2534
;  INFORMATION FOR SEQ ID NO:  6:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  240 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  oligodeoxynucleotide
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  YES
US-08-628-417-6

  Query Match             2.9%;  Score 117.8;  DB 1;  Length 240;
  Best Local Similarity   91.2%;  Pred. No. 4.8e-16;
  Matches  125;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       3928 CAATAAAGTTTCTGTCTCACTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3987
             ||||||| ||  |    | ||| |  ||||||||||||||||||||||||||||||||||||
Db         13 CAATAAAACTTTAGAAATAATTTTACTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 72

Qy       3988 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4047
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 132

Qy       4048 AAAAAAAAAAAAAAAA 4064
             ||||||||||||||||
Db        133 AAAAAAAAAAAAAAAA 149


RESULT 8
US-08-702-344-26
; Sequence 26, Application US/08702344
; Patent No. 5723315
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John
;    APPLICANT:  LaVallie, Edward
;    APPLICANT:  Racie, Lisa
;    APPLICANT:  Merberg, David
;    APPLICANT:  Treacy, Maurice
;    APPLICANT:  Spaulding, Vikki
;    TITLE OF INVENTION:  SECRETED PROTEINS AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  ENCODING THEM
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/702,344
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  144 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-702-344-26

  Query Match             2.9%;  Score 117.6;  DB 1;  Length 144;
  Best Local Similarity   90.4%;  Pred. No. 4.4e-16;
  Matches  123;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       3929 AATAAAGTTTCTGTCTCACTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3988
             || |||          |   ||||||||||||||||||||||||||||||||||||||||||
Db          1 AAAAAAAAAANNTNNTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 60

Qy       3989 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4048
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 120

Qy       4049 AAAAAAAAAAAAAAAA 4064
             ||||||||||||||||
Db        121 AAAAAAAAAAAAAAAA 136


RESULT 9
US-09-073-569-1
; Sequence 1, Application US/09073569
; Patent No. 6084088
;  GENERAL INFORMATION:
;    APPLICANT:  Sheppard, Paul O.
;    APPLICANT:  Grossmann, Angelika
;    TITLE OF INVENTION:  NOVEL TUMOR ANTIGENS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,569
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sawislak, Deborah A
;      REGISTRATION NUMBER:  37,438
;      REFERENCE/DOCKET NUMBER:  97-14
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6672
;      TELEFAX:  206-442-6678
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1733 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  34...1344
;      OTHER INFORMATION:
US-09-073-569-1

  Query Match             2.9%;  Score 117.4;  DB 3;  Length 1733;
  Best Local Similarity   95.3%;  Pred. No. 1.2e-15;
  Matches  121;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       3937 TTCTGTCTCACTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3996
             | || |||| |   ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1597 TCCTTTCTCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1656
```

```
Qy       3997 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4056
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1657 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1716

Qy       4057 AAAAAAA 4063
             |||||||
Db       1717 AAAAAAA 1723


RESULT 10
US-09-247-373B-33
; Sequence 33, Application US/09247373B
; Patent No. 6168954
; GENERAL INFORMATION:
;  APPLICANT:  MCGONIGLE, BRIAN
;  APPLICANT:  O'KEEFE, DANIEL
;  TITLE OF INVENTION:  SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
;  FILE REFERENCE:  CL-1108-A
;  CURRENT APPLICATION NUMBER: US/09/247,373B
;  CURRENT FILING DATE:  1999-02-10
;  PRIOR APPLICATION NUMBER:  08/924,747
;  PRIOR FILING DATE:  1997-09-05
;  NUMBER OF SEQ ID NOS:  56
;  SOFTWARE:  Microsoft Office 97
; SEQ ID NO 33
;   LENGTH: 1117
;   TYPE: DNA
;   ORGANISM: SOYBEAN
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1101)
;   OTHER INFORMATION: M=A OR C
;   NAME/KEY: unsure
;   LOCATION: (1104)
;   OTHER INFORMATION: M=A OR C
;   NAME/KEY: unsure
;   LOCATION: (1116)
;   OTHER INFORMATION: N=G or A or T or C
US-09-247-373B-33

  Query Match             2.8%;  Score 115.6;  DB 4;  Length 1117;
  Best Local Similarity   97.5%;  Pred. No. 2.5e-15;
  Matches  115;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3947 CTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4006
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        993 CTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1052

Qy       4007 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4064
             |||||||||||||||||||||||||||||||||||| ||||||:||:||||||
Db       1053 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAAAAAMAAMAAAAAA 1110


RESULT 11
US-09-363-708-3
; Sequence 3, Application US/09363708
; Patent No. 6399747
;  GENERAL INFORMATION:
;    APPLICANT:  Schmandt, et al.
;    TITLE OF INVENTION:  NOVEL SHC BINDING PROTEIN
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  233 South Wacker Drive/6300 Sears Tower
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/363,708
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  01017/34451
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2246 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;      DESCRIPTION:  /desc = "mouse PAL cDNA"
US-09-363-708-3

  Query Match             2.8%;  Score 114.2;  DB 4;  Length 2246;
  Best Local Similarity   97.5%;  Pred. No. 6.5e-15;
  Matches  116;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3946 ACTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4005
             |||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2127 ACTTGACCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2186

Qy       4006 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4064
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2187 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2245


RESULT 12
US-08-702-344-3
; Sequence 3, Application US/08702344
; Patent No. 5723315
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John
;    APPLICANT:  LaVallie, Edward
;    APPLICANT:  Racie, Lisa
;    APPLICANT:  Merberg, David
;    APPLICANT:  Treacy, Maurice
;    APPLICANT:  Spaulding, Vikki
;    TITLE OF INVENTION:  SECRETED PROTEINS AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  ENCODING THEM
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/702,344
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  117 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-702-344-3

  Query Match             2.8%;  Score 114;  DB 1;  Length 117;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-15;
  Matches  114;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3951 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4010
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 60

Qy       4011 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4064
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 114


RESULT 13
US-09-245-041-10
; Sequence 10, Application US/09245041
; Patent No. 6274339
; GENERAL INFORMATION:
;  APPLICANT: Moore, K.
;  APPLICANT: Nagle, D.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT
;  TITLE OF INVENTION: OF BODY WEIGHT DISORDERS INCLUDING OBESITY
;  FILE REFERENCE: 7853-136
;  CURRENT APPLICATION NUMBER: US/09/245,041
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/093,630
;  EARLIER FILING DATE: 1998-07-21
;  EARLIER APPLICATION NUMBER: 60/104,978
;  EARLIER FILING DATE: 1998-10-20
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 10
;   LENGTH: 1051
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-245-041-10

  Query Match             2.8%;  Score 114;  DB 4;  Length 1051;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-15;
  Matches  114;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3951 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4010
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         10 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 69

Qy       4011 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4064
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 123


RESULT 14
US-09-443-041A-27
; Sequence 27, Application US/09443041A
; Patent No. 6465717
; GENERAL INFORMATION:
;  APPLICANT: Famodu, Omolayo O.
;  APPLICANT:  Orozco, Buddy
;  APPLICANT:  Rafalski, Antoni
;  APPLICANT:  Shen, Jennie
;  TITLE OF INVENTION: Sterol Metabolism Enzymes
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/09/443,041A
;  CURRENT FILING DATE:  1999-11-18
;  PRIOR APPLICATION NUMBER: 60/109,283
;  PRIOR FILING DATE: 1998-11-20
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 27
;   LENGTH: 1447
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-443-041A-27

  Query Match             2.8%;  Score 114;  DB 4;  Length 1447;
  Best Local Similarity   100.0%;  Pred. No. 6e-15;
  Matches  114;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3951 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4010
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1318 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1377

Qy       4011 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4064
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1378 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1431


RESULT 15
US-09-394-645-1
; Sequence 1, Application US/09394645
; Patent No. 6380371
; GENERAL INFORMATION:
;  APPLICANT: Sassetti, Christopher M.
;  APPLICANT: Rosen, Steven R.
;  TITLE OF INVENTION: Endoglycan: A NOVEL PROTEIN HAVING SELECTIN
;  TITLE OF INVENTION: LIGAND AND CHEMOKINE PRESENTATION ACTIVITY
;  FILE REFERENCE: 6510-122US1
;  CURRENT APPLICATION NUMBER: US/09/394,645
;  CURRENT FILING DATE: 1999-09-13
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 2269
;   TYPE: DNA
;   ORGANISM: human
US-09-394-645-1

  Query Match             2.8%;  Score 114;  DB 4;  Length 2269;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-15;
  Matches  114;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3951 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4010
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2156 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2215

Qy       4011 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4064
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2216 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2269


Search completed: July 22, 2003, 06:25:08
Job time : 167.893 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2003, 18:26:42 ; Search time 784.537 Seconds
(without alignments)
10686.606 Million cell updates/sec

Title: US-09-873-737A-3
Perfect score: 4064
Sequence: 1 agaagcctgctacagggggt.....aaaaaaaaaaaaaaaaaaa 4064

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
11: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq3:*
13: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4059.6	99.9	4064	10	US-09-873-737A-3	Sequence 3, Appli
	2	1833.6	45.1	3472	10	US-09-873-737A-5	Sequence 5, Appli
	3	1680.6	41.4	2328	14	US-10-043-774B-1	Sequence 1, Appli
	4	1164.2	28.6	2292	11	US-09-998-598-2602	Sequence 2602, Ap
	5	475.6	11.7	2009	15	US-10-029-495-8	Sequence 8, Appli
	6	399.4	9.8	3649	10	US-09-801-574-15	Sequence 15, Appl
	7	311.8	7.7	423	15	US-10-102-806-267	Sequence 267, App
	8	311.4	7.7	3047	10	US-09-873-737A-1	Sequence 1, Appli
	9	183.4	4.5	560	15	US-10-029-495-1	Sequence 1, Appli
	10	144.8	3.6	889	15	US-10-027-632-166243	Sequence 166243,
	11	144.8	3.6	889	15	US-10-027-632-166244	Sequence 166244,
	12	144.8	3.6	889	15	US-10-027-632-166245	Sequence 166245,
	13	125	3.1	444	11	US-09-960-352-1281	Sequence 1281, Ap
	14	123.8	3.0	1225	10	US-09-908-711-53	Sequence 53, Appl
	15	123.8	3.0	2288	15	US-10-106-698-464	Sequence 464, App
	16	123.2	3.0	1696	15	US-10-274-971-1	Sequence 1, Appli
c	17	122	3.0	396	10	US-09-825-294-16	Sequence 16, Appl
c	18	122	3.0	396	11	US-09-970-966-16	Sequence 16, Appl
c	19	122	3.0	396	14	US-10-212-677-16	Sequence 16, Appl
c	20	120.6	3.0	15832	15	US-10-239-676-106	Sequence 106, App
c	21	120.4	3.0	611	15	US-10-198-846-8503	Sequence 8503, Ap
	22	120.2	3.0	418	12	US-09-918-995-7350	Sequence 7350, Ap
	23	120.2	3.0	3466	10	US-09-726-643-35	Sequence 35, Appl
	24	120.2	3.0	3466	15	US-10-042-141-35	Sequence 35, Appl
c	25	120	3.0	286	11	US-09-960-352-13342	Sequence 13342, A
	26	120	3.0	464	15	US-10-106-698-307	Sequence 307, App
	27	119.8	2.9	327	11	US-09-960-352-4630	Sequence 4630, Ap
	28	119.6	2.9	900	15	US-10-076-785-47	Sequence 47, Appl
	29	119.6	2.9	2645	10	US-09-764-870-76	Sequence 76, Appl
	30	119.6	2.9	2645	15	US-10-125-540-76	Sequence 76, Appl
	31	119.4	2.9	1457	12	US-09-948-820-38	Sequence 38, Appl
	32	119.2	2.9	429	11	US-09-960-352-11097	Sequence 11097, A
	33	119.2	2.9	3010	15	US-10-227-884-7	Sequence 7, Appli
	34	119.2	2.9	3010	15	US-10-230-163-7	Sequence 7, Appli
	35	119.2	2.9	3010	15	US-10-230-338-7	Sequence 7, Appli
	36	119.2	2.9	3010	15	US-10-218-631-7	Sequence 7, Appli
	37	119.2	2.9	3010	15	US-10-230-414-7	Sequence 7, Appli
	38	119.2	2.9	3010	15	US-10-216-159A-7	Sequence 7, Appli
	39	119.2	2.9	3010	15	US-10-218-849-7	Sequence 7, Appli
	40	119.2	2.9	3010	15	US-10-227-873-7	Sequence 7, Appli
	41	119.2	2.9	3010	15	US-10-227-883-7	Sequence 7, Appli
	42	119.2	2.9	3010	15	US-10-219-076-7	Sequence 7, Appli
	43	119.2	2.9	3010	15	US-10-230-434-7	Sequence 7, Appli
	44	119.2	2.9	3010	15	US-10-219-003-7	Sequence 7, Appli
	45	119.2	2.9	3010	15	US-10-219-075-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-873-737A-3
; Sequence 3, Application US/09873737A
; Patent No. US20020076797A1
; GENERAL INFORMATION:
; APPLICANT: Duke University
; APPLICANT: Lin, Haifan
; TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
; TITLE OF INVENTION: PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
; FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2
; CURRENT APPLICATION NUMBER: US/09/873,737A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/US99/28764
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/110,901
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 4064
; TYPE: DNA
; ORGANISM: Mus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (191)..(2776)
; NAME/KEY: misc_feature
; LOCATION: (458)
; OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (836)
; OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (1337)
; OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (2636)
; OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
US-09-873-737A-3

Query Match 99.9%; Score 4059.6; DB 10; Length 4064;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4064; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 AGAAGCCTGCTACAGGGGGTGGCGTGAGGAGCCTCAAGGACTGCGGACCTAAGCGTGCTA 60
Db       1 AGAAGCCTGCTACAGGGGGTGGCGTGAGGAGCCTCAAGGACTGCGGACCTAAGCGTGCTA 60

Qy      61 AGCACGGGGCCAGGATAGGACTGGACGCTGCTGGACCACGACGATCAGGGAGTGACCCCC 120
Db      61 AGCACGGGGCCAGGATAGGACTGGACGCTGCTGGACCACGACGATCAGGGAGTGACCCCC 120

Qy     121 ACCAGCATCCAGCCTGCGGCAGTTGTAGGACTGGAAGGAGGCCAAGCTGAACACCTGAGC 180
Db     121 ACCAGCATCCAGCCTGCGGCAGTTGTAGGACTGGAAGGAGGCCAAGCTGAACACCTGAGC 180

Qy     181 TGACTGGAAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGA 240
Db     181 TGACTGGAAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGA 240

Qy     241 GACGGTGCAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACC 300
Db     241 GACGGTGCAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACC 300

Qy     301 TCAACAGTCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGT 360
Db     301 TCAACAGTCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGT 360

Qy     361 CGGAGCCACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACT 420
Db     361 CGGAGCCACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACT 420

Qy     421 GGCAGAGAGAGGAGGGCGTCGCCGAGACTTCCATGACNTTGGTGTGAACACCAGACAGAA 480
Db     421 GGCAGAGAGAGGAGGGCGTCGCCGAGACTTCCATGACNTTGGTGTGAACACCAGACAGAA 480

Qy     481 CCTTGACCATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCAC 540
Db     481 CCTTGACCATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCAC 540

Qy     541 CAACCACTTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTA 600
Db     541 CAACCACTTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTA 600

Qy     601 CAATCCCCTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCT 660
Db     601 CAATCCCCTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCT 660

Qy     661 CATTGGAAGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCA 720
Db     661 CATTGGAAGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCA 720

Qy     721 CAAGGTCACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCAC 780
Db     721 CAAGGTCACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCAC 780

Qy     781 CCTGACCAACGAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTT 840
Db     781 CCTGACCAACGAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTT 840

Qy     841 CAGGAGGCTCTTGAAAATCATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAG 900
Db     841 CAGGAGGCTCTTGAAAATCATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAG 900

Qy     901 TGACCCGATTGATATTCCAAACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCAT 960
Db     901 TGACCCGATTGATATTCCAAACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCAT 960

Qy     961 CCTTCAGTATGAGAACAACATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAG 1020
Db     961 CCTTCAGTATGAGAACAACATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAG 1020

Qy    1021 CGAGACTGTCCTAGACTTCATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCA 1080
Db    1021 CGAGACTGTCCTAGACTTCATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCA 1080

Qy    1081 GGAGCAAGTGTCGAAGGAGCTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGAC 1140
Db    1081 GGAGCAAGTGTCGAAGGAGCTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGAC 1140

Qy    1141 CTACCGGGTGGATGACATTGACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGA 1200
Db    1141 CTACCGGGTGGATGACATTGACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGA 1200

Qy    1201 TGGCTCGGAGGTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGA 1260
Db    1201 TGGCTCGGAGGTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGA 1260

Qy    1261 CCTGAAGCAGCCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCT 1320
Db    1261 CCTGAAGCAGCCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCT 1320

Qy    1321 GCCTGGCCCAGCTATGNTCATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAAT 1380
Db    1321 GCCTGGCCCAGCTATGNTCATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAAT 1380

Qy    1381 GCGCAATGATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCA 1440
Db    1381 GCGCAATGATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCA 1440

Qy    1441 GCGGCAGCGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAG 1500
Db    1441 GCGGCAGCGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAG 1500

Qy    1501 AGAGCTTCGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAAT 1560
Db    1501 AGAGCTTCGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAAT 1560

Qy    1561 CTTACAATCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGC 1620
Db    1561 CTTACAATCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGC 1620

Qy    1621 AGACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTG 1680
Db    1621 AGACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTG 1680

Qy    1681 GCTGCTGATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTT 1740
Db    1681 GCTGCTGATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTT 1740

Qy    1741 CAAAGTGACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGA 1800
Db    1741 CAAAGTGACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGA 1800

Qy    1801 CAGAACAGAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGT 1860
Db    1801 CAGAACAGAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGT 1860

Qy    1861 TGTCTGTCTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTG 1920
Db    1861 TGTCTGTCTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTG 1920

Qy    1921 TACAGACTGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAAC 1980
Db    1921 TACAGACTGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAAC 1980

Qy    1981 AGTCATGGCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCT 2040
Db    1981 AGTCATGGCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCT 2040

Qy    2041 CTGGCGGGTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGA 2100
Db    2041 CTGGCGGGTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGA 2100

Qy    2101 CACCACAGCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGAC 2160
```

```
Db       2101 CACCACAGCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGAC 2160
Qy       2161 CCGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAA 2220
Db       2161 CCGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAA 2220
Qy       2221 GGTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCG 2280
Db       2221 GGTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCG 2280
Qy       2281 TGTCATCGTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGA 2340
Db       2281 TGTCATCGTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGA 2340
Qy       2341 GGTCCCACAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGAC 2400
Db       2341 GGTCCCACAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGAC 2400
Qy       2401 TGTAATCGTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACT 2460
Db       2401 TGTAATCGTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACT 2460
Qy       2461 TCAGAACCCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGA 2520
Db       2461 TCAGAACCCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGA 2520
Qy       2521 CTTTTTCATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGT 2580
Db       2521 CTTTTTCATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGT 2580
Qy       2581 CATCTATGACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCTG 2640
Db       2581 CATCTATGACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCTG 2640
Qy       2641 CCACGTGTACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACA 2700
Db       2641 CCACGTGTACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACA 2700
Qy       2701 CAAGCTGGCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAA 2760
Db       2701 CAAGCTGGCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAA 2760
Qy       2761 CCGCCTCTACTACCTCTAACCAGCAGCGAGGGCCGGCATGCTGGTTTCCTTCAGAGCTCT 2820
Db       2761 CCGCCTCTACTACCTCTAACCAGCAGCGAGGGCCGGCATGCTGGTTTCCTTCAGAGCTCT 2820
Qy       2821 GCCCAATCCTTCAGTTCTATTTTTCATAGGATGAGGTTCTAGCGGGATGTTGGGGGTAGG 2880
Db       2821 GCCCAATCCTTCAGTTCTATTTTTCATAGGATGAGGTTCTAGCGGGATGTTGGGGGTAGG 2880
Qy       2881 GGTTGGGGGCCGCTGTTTTGTTTTGTTTTGTTTTTAGTTGTTGGTGTGTAGCGGCTATAG 2940
Db       2881 GGTTGGGGGCCGCTGTTTTGTTTTGTTTTGTTTTTAGTTGTTGGTGTGTAGCGGCTATAG 2940
Qy       2941 AGTTAACTAAGGCTTATGTTTGCTTTGTTTCTCATGGTGTGTTGTGACAGTTTGCTTGGC 3000
Db       2941 AGTTAACTAAGGCTTATGTTTGCTTTGTTTCTCATGGTGTGTTGTGACAGTTTGCTTGGC 3000
Qy       3001 AGAGTGGTGGCAAGGGCCGTGGTACCACAGATAGACTTTCCTAAAGCAGCCGTGTTGGCA 3060
Db       3001 AGAGTGGTGGCAAGGGCCGTGGTACCACAGATAGACTTTCCTAAAGCAGCCGTGTTGGCA 3060
Qy       3061 GGTTCTTGGAGGTGCTCCTGGGAGCACCTCCTATTGTACGCTGCTCTCTCTCTCTCTCTC 3120
Db       3061 GGTTCTTGGAGGTGCTCCTGGGAGCACCTCCTATTGTACGCTGCTCTCTCTCTCTCTCTC 3120
Db       3121 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGCATGTTTGAATTTAAAAGAGGA 3180
Qy       3121 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGCATGTTTGAATTTAAAAGAGGA 3180
Qy       3181 AAGGACACAAGGCAAGAATGCCCAGAGCATCCTAGATGGGTGCTATATGGGACTCTGACA 3240
Db       3181 AAGGACACAAGGCAAGAATGCCCAGAGCATCCTAGATGGGTGCTATATGGGACTCTGACA 3240
Qy       3241 CCTCAGGTTCTTAAAGATAAAAAGGCCAGAGGTACCTTTCTGTAATTCTTTTAAGAGTTA 3300
Db       3241 CCTCAGGTTCTTAAAGATAAAAAGGCCAGAGGTACCTTTCTGTAATTCTTTTAAGAGTTA 3300
Qy       3301 GGTCATAAGGAATGAAAAGAGTTACTTTTCTCCCTCCATGAAAGTGTTTTTTAAGACTGA 3360
Db       3301 GGTCATAAGGAATGAAAAGAGTTACTTTTCTCCCTCCATGAAAGTGTTTTTTAAGACTGA 3360
Qy       3361 CAAGGCCGGTGGTGGAGTGAGTGTTGTTGTTTTGTTTCGCAATTTTCTTACCAAGTGAAA 3420
Db       3361 CAAGGCCGGTGGTGGAGTGAGTGTTGTTGTTTTGTTTCGCAATTTTCTTACCAAGTGAAA 3420
Qy       3421 TTACTGACCATAATTTTATTTAATAGTTTTTAGTTTTTGAAAGCCTGTGTTTTCTTTGCT 3480
Db       3421 TTACTGACCATAATTTTATTTAATAGTTTTTAGTTTTTGAAAGCCTGTGTTTTCTTTGCT 3480
Qy       3481 TTTAAACTACAGGCGACCACTCTGGGATAGGATTCCACTCCCAAAAGTAATGTGTTGGGT 3540
Db       3481 TTTAAACTACAGGCGACCACTCTGGGATAGGATTCCACTCCCAAAAGTAATGTGTTGGGT 3540
Qy       3541 GTGGTGGCACACGCCTGTGATCTCAGGACTCAGGAGGTGAAGCCAGGGGAACCCCAAGTT 3600
Db       3541 GTGGTGGCACACGCCTGTGATCTCAGGACTCAGGAGGTGAAGCCAGGGGAACCCCAAGTT 3600
Qy       3601 CCCGGTCAGCCTAGGCTACATAGTGAGACTTTGTCTTAACAAACAGAAAAAAAGAAAAAG 3660
Db       3601 CCCGGTCAGCCTAGGCTACATAGTGAGACTTTGTCTTAACAAACAGAAAAAAAGAAAAAG 3660
Qy       3661 AAAGAAAAAAAAAAAGAAAGAAAAAAAAAARAAAGAAAAATGGAGCAGTATGAGTTATGGGT 3720
Db       3661 AAAGAAAAAAAAAAAGAAAGAAAAAAAAAARAAAGAAAAATGGAGCAGTATGAGTTATGGGT 3720
Qy       3721 GCCAGTTTCTGCAGGAGACTTTGGGAGTAAGAGCTGCCTCTGTTCCTAAGAGGCTGCAAG 3780
Db       3721 GCCAGTTTCTGCAGGAGACTTTGGGAGTAAGAGCTGCCTCTGTTCCTAAGAGGCTGCAAG 3780
Qy       3781 GCTGCGGCTGGGGGGCCAGTTCAGGTGCCAGGTGAGGTACGGCCACATCTATTTGCCTGG 3840
Db       3781 GCTGCGGCTGGGGGGCCAGTTCAGGTGCCAGGTGAGGTACGGCCACATCTATTTGCCTGG 3840
Qy       3841 CTCTTTGACAAAGAGAGCAGCTTAGAGTGGCTAGAGTGTGATGAGTTTTACACAATGGCG 3900
Db       3841 CTCTTTGACAAAGAGAGCAGCTTAGAGTGGCTAGAGTGTGATGAGTTTTACACAATGGCG 3900
Qy       3901 ATGTGTGTGTGTGCCCCTGGCACAGCTCAATAAAGTTTCTGTCTCACTTTAAAAAAAAAA 3960
Db       3901 ATGTGTGTGTGTGCCCCTGGCACAGCTCAATAAAGTTTCTGTCTCACTTTAAAAAAAAAA 3960
Qy       3961 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4020
Db       3961 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4020
Qy       4021 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4064
Db       4021 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4064
```

RESULT 2
US-09-873-737A-5
; Sequence 5, Application US/09873737A
; Patent No. US20020076797A1
; GENERAL INFORMATION:
; APPLICANT: Duke University
; APPLICANT: Lin, Haifan
; TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
; TITLE OF INVENTION: PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
; FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2
; CURRENT APPLICATION NUMBER: US/09/873,737A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/US99/28764
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/110,901

```
;  PRIOR FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 3472
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (167)..(2749)
;   NAME/KEY: misc_feature
;   LOCATION: (392)
;   OTHER INFORMATION: n=t or a, Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (1073)
;   OTHER INFORMATION: n=c or a, Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (2369)
;   OTHER INFORMATION: n=c or a, Xaa=Leu or Ile
US-09-873-737A-5

  Query Match             45.1%;  Score 1833.6;  DB 10;  Length 3472;
  Best Local Similarity   81.8%;  Pred. No. 0;
  Matches 2125;  Conservative    0;  Mismatches  471;  Indels    3;  Gaps    1;

Qy       188 AAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGAGACGGTG 247
Db       164 ACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCG 223

Qy       248 CAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACCTCAACAG 307
Db       224 CAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCG 283

Qy       308 TCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGTCGGAGCC 367
Db       284 CCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAA---CAGCAGGAGGA 340

Qy       368 ACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACTGGCAGAG 427
Db       341 ACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAG 400

Qy       428 AGAGGAGGGCGTCGCCGAGACTTCCATGACNTTGGTGTGAACACCAGACAGAACCTTGAC 487
Db       401 AGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGAC 460

Qy       488 CATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCAC 547
Db       461 CATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCAT 520

Qy       548 TTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCC 607
Db       521 TTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCA 580

Qy       608 CTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGA 667
Db       581 CTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGA 640

Qy       668 AGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGTC 727
Db       641 AAGTGTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTT 700

Qy       728 ACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCACCCTGACC 787
Db       701 ACTGAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACA 760

Qy       788 AACGAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTTCAGGAGG 847
Db       761 AATGAACTTCCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGG 820

Qy       848 CTCTTGAAAATCATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAGTGACCCG 907
Db       821 CTTTTGAAAATCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCA 880

Qy       908 ATTGATATTCCAAACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCATCCTTCAG 967
Db       881 ATTGATATTCCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAG 940

Qy       968 TATGAGAACAACATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAGCGAGACT 1027
Db       941 TATGAAAACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACT 1000

Qy      1028 GTCCTAGACTTCATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCAGGAGCAA 1087
Db      1001 GTTTTGGATTTCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAA 1060

Qy      1088 GTGTCGAAGGAGCTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGACCTACCGG 1147
Db      1061 GTTTCCAAAGAANTAATAGGTTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGA 1120

Qy      1148 GTGGATGACATTGACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGATGGCTCG 1207
Db      1121 GTGGATGATATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCT 1180

Qy      1208 GAGGTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGACCTGAAG 1267
Db      1181 GGGGTCAGCTTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAG 1240

Qy      1268 CAGCCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCTGCCTGGC 1327
Db      1241 CAGCCTGTCTTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGG 1300

Qy      1328 CCAGCTATGNTCATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAATGCGCAAT 1387
Db      1301 CCTGCCATGCTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAAT 1360

Qy      1388 GATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCAGCGGCAG 1447
Db      1361 GATTTTAACGTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAG 1420

Qy      1448 CGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAGAGAGCTT 1507
Db      1421 CGTGAAGTGGGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTT 1480

Qy      1508 CGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAATCTTACAA 1567
Db      1481 CGAGACTGGGGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAA 1540

Qy      1568 TCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGCAGACTGG 1627
Db      1541 ACAGAAAAGATTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGG 1600

Qy      1628 TCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGGCTGCTG 1687
Db      1601 TCCAAAGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTG 1660

Qy      1688 ATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTTCAAAGTG 1747
Db      1661 ATCTATACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTT 1720

Qy      1748 ACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGACAGAACA 1807
Db      1721 ACACCAGCCATGGGCATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACT 1780

Qy      1808 GAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGT 1867
Db      1781 GAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGT 1840

Qy      1868 CTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTGTACAGAC 1927
Db      1841 CTGTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGAT 1900

Qy      1928 TGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATG 1987
Db      1901 TGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATG 1960

Qy      1988 GCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGG 2047
```

```
Db       1961 GCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGG 2020

Qy       2048 GTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACACCACA 2107

Db       2021 GTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACA 2080

Qy       2108 GCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGG 2167

Db       2081 GCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGG 2140

Qy       2168 TTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGC 2227

Db       2141 TTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGC 2200

Qy       2228 TTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATC 2287

Db       2201 CTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATC 2260

Qy       2288 GTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCA 2347

Db       2261 GTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCA 2320

Qy       2348 CAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATC 2407

Db       2321 CAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATT 2380

Qy       2408 GTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAAC 2467

Db       2381 GTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAAT 2440

Qy       2468 CCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTC 2527

Db       2441 CCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTT 2500

Qy       2528 ATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTAT 2587

Db       2501 ATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTAT 2560

Qy       2588 GACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTG 2647

Db       2561 GACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATC 2620

Qy       2648 TACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTG 2707

Db       2621 TATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTG 2680

Qy       2708 GCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTC 2767

Db       2681 GCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTT 2740

Qy       2768 TACTACCTCTAACCAGCAG 2786

Db       2741 TACTACCTCTAACCTGCAG 2759
```

```
RESULT 3
US-10-043-774B-1
; Sequence 1, Application US/10043774B
; Publication No. US20030129599A1
; GENERAL INFORMATION:
;  APPLICANT: University of Illinois at Chicago
;  APPLICANT:  Sharma, Arun
;  APPLICANT:  Hoffman, Ronald
;  TITLE OF INVENTION: HUMAN HEMATOPOIETIC GROWTH REGULATORY GENE AND USES
;  FILE REFERENCE: MBHB: CU08/PPA
;  CURRENT APPLICATION NUMBER: US/10/043,774B
;  CURRENT FILING DATE:  2002-06-06
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 2328
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(2328)
;   OTHER INFORMATION: Human Hiwi Protein
US-10-043-774B-1

  Query Match             41.4%;  Score 1680.6;  DB 14;  Length 2328;
  Best Local Similarity   82.7%;  Pred. No. 2.7e-302;
  Matches 1920;  Conservative    0;  Mismatches  403;  Indels    0;  Gaps    0;
```

```
Qy        457 CNTTGGTGTGAACACCAGACAGAACCTTGACCATGTCAAAGAGTCAAAGACAGGCTCCTC 516

Db          6 CTTTGGTGTGAACACAAGGCAGAACTTAGACCATGTTAAAGAATCAAAAACAGGTTCTTC 65

Qy        517 TGGCATCATTGTGAAGCTGAGCACCAACCACTTCCGGCTGACCTCGCGCCCACAGTGGGC 576

Db         66 AGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGACATCCCGTCCCCAGTGGGC 125

Qy        577 CCTGTATCAGTACCACATCGACTACAATCCCCTGATGGAGGCCCGAAGGCTTCGCTCCGC 636

Db        126 CTTATATCAGTATCACATTGACTATAACCCACTGATGGAAGCCAGAAGACTCCGTTCAGC 185

Qy        637 ACTGCTCTTCCAGCATGAAGACCTCATTGGAAGGTGTCATGCTTTCGATGGGACAATATT 696

Db        186 TCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCATGCTTTTGATGGAACGATATT 245

Qy        697 GTTTTTACCTAAGAGACTACAGCACAAGGTCACAGAAGTATTCAGTCAGACTCGGAATGG 756

Db        246 ATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTTTTTAGTAAGACCCGGAATGG 305

Qy        757 GGAACACGTGAGGATCACCATCACCCTGACCAACGAGCTGCCGCCCACCTCGCCCACCTG 816

Db        306 AGAGGATGTGAGGATAACGATCACTTTAACAAATGAACTTCCACCTACATCACCAACTTG 365

Qy        817 CCTGCAGTTCTATAACATCNTCTTCAGGAGGCTCTTGAAAATCATGAATTTGCAACAAAT 876

Db        366 TTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAAATCATGAATTTGCAACAAAT 425

Qy        877 TGGACGGAATTATTACAATCCAAGTGACCCGATTGATATTCCAAACCACAGGTTGGTGAT 936

Db        426 TGGACGAAATTATTATAACCCAAATGACCCAATTGATATTCCAAGTCACAGGTTGGTGAT 485

Qy        937 CTGGCCCGGCTTCACCACCTCCATCCTTCAGTATGAGAACAACATCATGCTCTGCACAGA 996

Db        486 TTGGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAACAGCATCATGCTCTGCACTGA 545

Qy        997 CGTCAGCCACAAGGTGCTCCGCAGCGAGACTGTCCTAGACTTCATGTTCAATCTATACCA 1056

Db        546 CGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGATTTCATGTTCAACTTTTATCA 605

Qy       1057 GCAGACAGAGGAGCACAAGTTCCAGGAGCAAGTGTCGAAGGAGCTCATAGGCCTCATCGT 1116

Db        606 TCAGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAAGAACTAATAGGTTTAGTTGT 665

Qy       1117 TCTCACCAAGTACAATAACAAGACCTACCGGGTGGATGACATTGACTGGGACCAGAATCC 1176

Db        666 TCTTACCAAGTATAACAATAAGACATACAGAGTGGATGATATTGACTGGGACCAGAATCC 725

Qy       1177 AAAGAGCACCTTCAAGAAGGCGGATGGCTCGGAGGTCAGCTTCCTGGAGTACTACAGGAA 1236

Db        726 CAAGAGCACCTTTAAGAAAGCCGACGGCTCTGAAGTCAGCTTCTTAGAATACTACAGGAA 785

Qy       1237 GCAATACAACCAGGAGATCACGGACCTGAAGCAGCCGGTGCTGGTGAGCCAACCCAAGCG 1296

Db        786 GCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGCCAGCCCAAGAG 845

Qy       1297 GAGGAGAGGCCCCGGCGGCACCCTGCCTGGCCCAGCTATGNTCATCCCTGAACTCTGCTA 1356

Db        846 AAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATGCTCATTCCTGAGCTCTGCTA 905

Qy       1357 TCTCACAGGCCTGACTGATAAAATGCGCAATGATTTCAATGTGATGAAGGACCTGGCAGT 1416
```

```
Db        906 TCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAACGTGATGAAAGACTTAGCCGT 965

Qy       1417 GCACACGCGGCTGACCCCTGAGCAGCGGCAGCGGGAGGTGGGCCGCCTCATCGACTACAT 1476

Db        966 TCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTGGGACGACTCATTGATTACAT 1025

Qy       1477 CCACAAGGATGACAATGTGCAGAGAGAGCTTCGAGACTGGGGCCTGAGCTTCGACTCAAA 1536

Db       1026 TCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGGGTTTGAGCTTTGATTCCAA 1085

Qy       1537 CTTGCTGTCCTTCTCTGGAAGAATCTTACAATCTGAGAAGATCCACCAGGGCGGAAAGAC 1596

Db       1086 CTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAGATTCACCAAGGTGGAAAAAC 1145

Qy       1597 GTTTGATTACAACCCACAATTTGCAGACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGAT 1656

Db       1146 ATTTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAACAAGAGGTGCACCATTAAT 1205

Qy       1657 CAGCGTGAAGCCATTGGATAACTGGCTGCTGATCTATACCCGCAGGAATTATGAAGCAGC 1716

Db       1206 TAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACGCGAAGAAATTATGAAGCAGC 1265

Qy       1717 CAACTCACTGATACAGAACCTGTTCAAAGTGACTCCAGCCATGGGCATCCAGATGAAAAA 1776

Db       1266 CAATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGCATGCAAATGAGAAA 1325

Qy       1777 GGCAATCATGATCGAGGTGGATGACAGAACAGAAGCTTATCTGAGAGCCTTGCAGCAGAA 1836

Db       1326 AGCAATAATGATTGAAGTGGATGACAGAACTGAAGCCTACTTAAGAGTCTTACAGCAAAA 1385

Qy       1837 GGTGACGTCAGACACTCAGATAGTTGTCTGTCTCTTGTCAAGTAATCGGAAGGACAAATA 1896

Db       1386 GGTCACAGCAGACACCCAGATAGTTGTCTGTCTGTTGTCAAGTAATCGGAAGGACAAATA 1445

Qy       1897 TGATGCCATCAAGAAGTACTTGTGTACAGACTGCCCCACCCCAAGTCAGTGTGTGGTGGC 1956

Db       1446 CGATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGC 1505

Qy       1957 CCGGACCCTGGGCAAGCAGCAAACAGTCATGGCCATTGCCACCAAGATCGCCCTGCAGAT 2016

Db       1506 CCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAAAGATTGCCCTACAGAT 1565

Qy       2017 GAACTGCAAGATGGGAGGCGAGCTCTGGCGGGTGGACATGGCCCTGAAACTGGCAATGAT 2076

Db       1566 GAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCTCGTGATGAT 1625

Qy       2077 CGTGGGCATCGACTGTTACCATGACACCACAGCTGGGCGGAGGTCCATCGCAGGATTCGT 2136

Db       1626 CGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGCAGGATTTGT 1685

Qy       2137 CGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCGG 2196

Db       1686 TGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCATATTTCAGGATAGAGG 1745

Qy       2197 GCAGGAGCTGGTGGATGGTCTCAAGGTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGG 2256

Db       1746 ACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGGCTCTGAGGGCTTGGAATAG 1805

Qy       2257 CTGCAATGAATACATGCCCAGCCGTGTCATCGTGTACCGAGACGGTGTGGGGGACGGGCA 2316

Db       1806 CTGCAATGAGTACATGCCCAGCCGGATCATCGTGTACCGCGATGGCGTAGGAGACGGCCA 1865

Qy       2317 GCTGAAGACCCTGGTCAATTATGAGGTCCCACAGTTCCTAGATTGCCTCAAGTCAGTCGG 2376

Db       1866 GCTGAAAACACTGGTGAACTACGAAGTGCCACAGTTTTTGGATTGTCTAAAATCCATTGG 1925

Qy       2377 GAGAGGTTACAACCCAAGACTGACTGTAATCGTGGTGAAGAAGCGTGTCAATGCCAGGTT 2436

Db       1926 TAGAGGTTACAACCCTAGACTAACGGTAATTGTGGTGAAGAAAAGAGTGAACACCAGATT 1985

Qy       2437 TTTTGCTCAGTCTGGGGGAAGACTTCAGAACCCTCTTCCAGGGACAGTCATCGATGTGGA 2496

Db       1986 TTTTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTCCTGGAACAGTTATTGATGTAGA 2045

Qy       2497 AGTCACCAGACCAGAGTGGTATGACTTTTTCATCGTGAGCCAGGCAGTGAGAAGCGGGAG 2556

Db       2046 GGTTACCAGACCAGAATGGTATGACTTTTTTATCGTGAGCCAGGCTGTGAGAAGTGGTAG 2105

Qy       2557 TGTGTCCCCAACACACTACAATGTCATCTATGACAGCAGTGGCCTGAAGCCCGACCACAT 2616

Db       2106 TGTTTCTCCCACACATTACAATGTCATCTATGACAACAGCGGCCTGAAGCCAGACCACAT 2165

Qy       2617 CCAGCGGCTGACATACAAGNTCTGCCACGTGTACTATAATTGGCCTGGAGTCATCCGAGT 2676

Db       2166 ACAGCGCTTGACCTACAAGCTGTGCCACATCTATTACAACTGGCCAGGTGTCATTCGTGT 2225

Qy       2677 CCCTGCACCTTGCCAGTATGCACACAAGCTGGCCTTCCTCGTGGGCCAGAGCATCCACAG 2736

Db       2226 TCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTGTTGGCCAGAGTATTCACAG 2285

Qy       2737 AGAGCCAAACCTCTCCCTGTCCAACCGCCTCTACTACCTCTAA 2779

Db       2286 AGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAA 2328


RESULT 4
US-09-998-598-2602
; Sequence 2602, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
;  APPLICANT: Stolk, John A.
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Chenault, Ruth A.
;  APPLICANT:  Meagher, Madelein Joy
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.561
;  CURRENT APPLICATION NUMBER: US/09/998,598
;  CURRENT FILING DATE:  2001-11-16
;  NUMBER OF SEQ ID NOS: 2606
;  SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 2602
;   LENGTH: 2292
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-998-598-2602

  Query Match             28.6%;  Score 1164.2;  DB 11;  Length 2292;
  Best Local Similarity   83.5%;  Pred. No. 1.8e-206;
  Matches 1319;  Conservative    0;  Mismatches  260;  Indels    0;  Gaps    0;

Qy       1208 GAGGTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGACCTGAAG 1267

Db          1 GGGGTCAGCTTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAG 60

Qy       1268 CAGCCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCTGCCTGGC 1327

Db         61 CAGCCTGTCTTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGG 120

Qy       1328 CCAGCTATGNTCATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAATGCGCAAT 1387

Db        121 CCTGCCATGCTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAAT 180

Qy       1388 GATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCAGCGGCAG 1447

Db        181 GATTTTAACGTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAG 240

Qy       1448 CGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAGAGAGCTT 1507

Db        241 CGTGAAGTGGGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTT 300

Qy       1508 CGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAATCTTACAA 1567

Db        301 CGAGACTGGGGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAA 360

Qy       1568 TCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGCAGACTGG 1627
```

```
Db        361 ACAGAAAAGATTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGG 420

Qy       1628 TCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGGCTGCTG 1687

Db        421 TCCAAAGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTG 480

Qy       1688 ATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTTCAAAGTG 1747

Db        481 ATCTATACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTT 540

Qy       1748 ACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGACAGAACA 1807

Db        541 ACACCAGCCATGGGCATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACT 600

Qy       1808 GAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGT 1867

Db        601 GAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGT 660

Qy       1868 CTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTGTACAGAC 1927

Db        661 CTGTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGAT 720

Qy       1928 TGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATG 1987

Db        721 TGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATG 780

Qy       1988 GCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGG 2047

Db        781 GCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGG 840

Qy       2048 GTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACACCACA 2107

Db        841 GTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACA 900

Qy       2108 GCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGG 2167

Db        901 GCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGG 960

Qy       2168 TTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGC 2227

Db        961 TTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGC 1020

Qy       2228 TTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATC 2287

Db       1021 CTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATC 1080

Qy       2288 GTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCA 2347

Db       1081 GTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCA 1140

Qy       2348 CAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATC 2407

Db       1141 CAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGACTAACGGTAATT 1200

Qy       2408 GTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAAC 2467

Db       1201 GTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAAT 1260

Qy       2468 CCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTC 2527

Db       1261 CCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTT 1320

Qy       2528 ATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTAT 2587

Db       1321 ATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTAT 1380

Qy       2588 GACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTG 2647

Db       1381 GACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATC 1440

Qy       2648 TACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTG 2707

Db       1441 TATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTG 1500

Qy       2708 GCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTC 2767

Db       1501 GCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTT 1560

Qy       2768 TACTACCTCTAACCAGCAG 2786

Db       1561 TACTACCTCTAACCTGCAG 1579
```

RESULT 5
US-10-029-495-8
; Sequence 8, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
; APPLICANT: Drmanac, Radoje T.
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Dickson, Mark C.
; APPLICANT: Jones, Lee William
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Xue, Aidong
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi,Vinod
; TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
; FILE REFERENCE: 28110/35905A
; CURRENT APPLICATION NUMBER: US/10/029,495
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 09/378,667
; PRIOR FILING DATE: 1999-08-20
; PRIOR APPLICATION NUMBER: 09/687,527
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 2009
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (73)..(1569)
US-10-029-495-8

Query Match 11.7%; Score 475.6; DB 15; Length 2009;
Best Local Similarity 57.8%; Pred. No. 1.1e-78;
Matches 904; Conservative 0; Mismatches 651; Indels 9; Gaps 3;

```
Qy       1216 CTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGACCTGAAGCAGCCGGT 1275

Db         15 CTAGGTGGATTACTACAAGCAGCAGTATGATATTACTGTATCGGACCTGAATCAGCCCAT 74

Qy       1276 GCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCTGCCTGGCCCAGCTAT 1335

Db         75 GCTTGTTAGTCTGTTAAAGAAGAAGAGAAAT---GACAACAGTGAGCCTCAGCTCGCCCA 131

Qy       1336 GNTCATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAATGCGCAATGATTTCAA 1395

Db        132 CCTGATACCTGAGCTCTGCTTTCTAACAGGGCTGACTGACCAGGCAACATCTGATTTCCA 191

Qy       1396 TGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCAGCGGCAGCGGGAGGT 1455

Db        192 GCTGATGAAGGCTGTGGCTGAAAAGACACGTCTCAGTCCTTCAGGCCGGCAGCAGCGCCT 251

Qy       1456 GGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAGAGAGCTTCGAGACTG 1515

Db        252 GGCCAGGCTTGTGGACAACATCCAGAGGAATACCAATGCTCGCTTTGAACTAGAGACCTG 311

Qy       1516 GGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAATCTTACAATCTGAGAA 1575

Db        312 GGGACTGCATTTTG---GAAGCCAGATATCTCTGACTGGCCGGATTGTGCCTTCAGAAAA 368
```

```
Qy       1576 GATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGCAGACTGGTCCAAAGA 1635
Db        369 AATATTAATGCAAGACCACATATGTCAACCTGTGTCTG---CTGCTGACTGGTCCAAGGA 425

Qy       1636 AACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGGCTGCTGATCTATAC 1695
Db        426 TATTCGAACTTGCAAGATTTTAAATGCACAGTCTTTGAATACCTGGTTGATTTTATGTAG 485

Qy       1696 CCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTTCAAAGTGACTCCAGC 1755
Db        486 CGACAGAACTGAATATGTTGCCGAGAGCTTTCTGAACTGCTTGAGAAGAGTTGCAGGTTC 545

Qy       1756 CATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGACAGAACAGAAGCTTA 1815
Db        546 CATGGGATTTAATGTGGACTACCCCAAAATCATAAAAGTACAAGAAAATCCAGCTGCATT 605

Qy       1816 TCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGTCTCTTGTC 1875
Db        606 TGTTAGAGCTATACAGCAATATGTTGATCCTGATGTTCAGCTGGTAATGTGCATTCTGCC 665

Qy       1876 AAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTGTACAGACTGCCCCAC 1935
Db        666 TTCTAATCAGAAGACCTATTATGATTCCATTAAAAAATATTTGAGCTCAGACTGCCCAGT 725

Qy       1936 CCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATGGCCATTGC 1995
Db        726 CCCAAGCCAATGTGTGCTTGCTCGGACCTTGAATAAACAGGGCATGATGATGAGTATCGC 785

Qy       1996 CACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGGGTGGACAT 2055
Db        786 CACCAAGATCGCTATGCAGATGACTTGCAAGCTCGGAGGCGAGCTGTGGGCTGTGGAAAT 845

Qy       2056 GGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACACCACAGCTGGGCG 2115
Db        846 ACCTTTAAAGTCCCTGATGGTGGTCGGTATTGATGTCTGTAAAGATGCACTCAGCAAGGA 905

Qy       2116 GAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCCCG 2175
Db        906 CGTGATGGTTGTTGGATGCGTGGCCAGTGTTAACCCCAGAATCACCAGGTGGTTTTCCCG 965

Qy       2176 CTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGCTTGCAAGC 2235
Db        966 CTGTATCCTTCAGAGAACAATGACTGATGTTGCAGATTGCTTGAAAGTTTTCATGACTGG 1025

Qy       2236 TGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATCGTGTACCG 2295
Db        1026 AGCACTCAACAAATGGTACAAGTACAATCATGATTTGCCAGCACGGATAATTGTGTACCG 1085

Qy       2296 AGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCACAGTTCCT 2355
Db        1086 TGCTGGTGTAGGGGATGGTCAGCTGAAAACACTTATTGAATATGAAGTCCCACAGCTGCT 1145

Qy       2356 AGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATCGTGGTGAA 2415
Db        1146 GAGCAGTGTGGCAGAATCCAGCTCAAATACCAGCTCAAGACTGTCGGTGATTGTGGTCAG 1205

Qy       2416 GAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAACCCTCTTCC 2475
Db        1206 GAAGAAGTGCATGCCACGATTCTTTACCGAAATGAACCGCACTGTACAGAACCCCCCACT 1265

Qy       2476 AGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTCATCGTGAG 2535
Db        1266 TGGCACTGTTGTGGATTCAGAAGCAACACGTAACGAATGGTATGACTTTTATCTGATCAG 1325

Qy       2536 CCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTATGACAGCAG 2595
Db        1326 CCAGGTGGCCTGCCGGGGAACTGTTAGTCCTACCTACTATAATGTCATCTATGATGACAA 1385

Qy       2596 TGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTGTACTATAA 2655
Db        1386 CGGCTTGAAGCCCGACCATATGCAGAGACTTACATTCAAATTGTGCCACCTGTACTACAA 1445

Qy       2656 TTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTGGCCTTCCT 2715
```

```
Db        1446 CTGGCCGGGCATAGTCAGTGTCCCAGCACCATGTCAGTATGCTCACAAGCTGACCTTTCT 1505

Qy       2716 CGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTCTACTACCT 2775
Db        1506 GGGGGCACAAAGCATTCATAAAGAACCCAGTCTGGAATTAGCCAACCATCTCTTCTACCT 1565

Qy       2776 CTAA 2779
Db        1566 GTGA 1569
```

RESULT 6
US-09-801-574-15
; Sequence 15, Application US/09801574
; Patent No. US20020081592A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Peijing Jeremy
; APPLICANT: Page, David C.
; TITLE OF INVENTION: Reproduction-Specific Genes
; FILE REFERENCE: 0399.2007-002
; CURRENT APPLICATION NUMBER: US/09/801,574
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: 60/187,518
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: 60/261,557
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 3649
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-801-574-15

Query Match 9.8%; Score 399.4; DB 10; Length 3649;
Best Local Similarity 53.1%; Pred. No. 1.9e-64;
Matches 991; Conservative 0; Mismatches 853; Indels 21; Gaps 6;

```
Qy        941 CCCGGCTTCACCACCTCCATCCTTCAGTATGAGAACAACATCATGCTCTGCACAGACGTC 1000
Db          1 CCTGGCTATGCGGCTAGTATCCGGAGGACAGACGGGGGTCTCTTCCTGCTCGCTGATGTC 60

Qy       1001 AGCCACAAGGTGCTCCGCAGCGAGACTGTCCTAGACTTCATGTTCAATCTATACCAGCAG 1060
Db         61 TCTCATAAGGTCATTCGGAACGACTCTGTGCTGGATGTCATGCATGCTATCTACCAGCAG 120

Qy       1061 ACAGAGGAGCACAAGTTCCAGGAGCAAGTGTCGAAGGAGCTCATAGGCCTCATCGTTCTC 1120
Db         121 AACAAGGAGCAC---TTCCAGGACGAGTGCAGCAAGCTTCTGGTTGGCAGCATTGTCATC 177

Qy       1121 ACCAAGTACAATAACAAGACCTACCGGGTGGATGACATTGACTGGGACCAGAATCCAAAG 1180
Db         178 ACGCGCTACAACAATCGTACCTACCGAATCGATGATGTGGACTGGAACAAGACCCCTAAA 237

Qy       1181 AGCACCTTCAAGAAGGCGGATGGCTCGGAGGTCAGCTTCCTGGAGTACTACAGGAAGCAA 1240
Db         238 GACAGCTTTGTCATGTCGGACGGGAAGGAAATCACATTCCTGGAATACTACAGCAAAAAC 297

Qy       1241 TACAACCAGGAGATCACGGACCTGAAGCAGCCGGTGCTGGTGAGCCAACCCA---AGCGG 1297
Db         298 TATGGGATCACAGTCAAGGAAGATGACCAGCCGCTGCTGATCCACCGGCCCAGTGAGAGA 357

Qy       1298 AGGAGAGGCCCCGGCGGCACCCTGCCTGGCCCAGCTATGNTCATCCCTGAACTCTGCTAT 1357
Db         358 CAGAATAACCATGGCATGTTGCTGAAGGGCGAGATCCTGCTGCTGCCCGAGCTCTCCTTC 417

Qy       1358 CTCACAGGCCTGACTGATAAAATGCGCAATGATTTCAATGTGATGAAGGACCTGGCAGTG 1417
Db         418 ATGACGGGGATCCCTGAGAAGATGAAGAAGGACTTCAGGGCCATGAAGGACTTGACTCAG 477

Qy       1418 CACACGCGGCTGACCCCTGAGCAGCGGCAGCGGGAGGTGGGCCGCCTCATCGACTACATC 1477
```

```
Db        478 CAGATTAACCTGAGCCCCAAGCAGCACCACGGTGCTTTGGAATGCCTGCTGCAGAGAATT 537

Qy       1478 CACAAGGATGACAATGTGCAGAGAGAGCTTCGAGACTGGGGCCTGAGCTTCGACTCAAAC 1537

Db        538 TCACAAAACGAGACAGCCAGCAATGAGCTGACCCGCTGGGGGCTCAGTCTGCATAAAGAT 597

Qy       1538 TTGCTGTCCTTCTCTGGAAGAATCTTACAATCTGAGAAGATCCACCAGGGCGGAAAGACG 1597

Db        598 GTCCACAAGATTGAAGGTCGGCTTCTGCCAATGGAGAGGATCAAC---TTAAGGAACACT 654

Qy       1598 TTTGATTACAACCCACAATTTGCAGACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGATC 1657

Db        655 TCATTTGTCACATCGGAGGGCCTGAACTGGGTTAAGGAAGTGACCAGAGATGCTTCCATT 714

Qy       1658 AGCGTGAAGCCATTGGATAACTGGCTGCTGATCTATACCCGCAGGAATTATGAAGCAGCC 1717

Db        715 CTAACTATTCCCATGCATTTCTGGGCACTCTTTTATCCAAAGAGAGCAATGGACCAAGCC 774

Qy       1718 AACTCACTGATACAGAACCTGTTCAAAGTGACTCCAGCCATGGGCATCCAGATGAAAAAG 1777

Db        775 AGAGAACTGGTTAACATGTTGGAAAAGATTGCCGGGCCCATTGGCATGCGCACAAGCCCC 834

Qy       1778 GCAATCATGATCGAGGT---GGATGACAGAACAGAAGCTTATCTGAGAGCCTTGCAG--- 1831

Db        835 CCAGCCTGGGTTGAGCTGAAGGATGACCGAATAGAGACCTATATCAGGACCATTCAGTCC 894

Qy       1832 ---CAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGTCTCTTGTCAAGTAATCGGAAG 1888

Db        895 TTACTGGGAGTTGAGGGGAAGATACAAATGGTCGTTTGCATCATCATGGGCACACGTGAT 954

Qy       1889 GACAAATATGATGCCATCAAGAAGTACTTGTGTACAGACTGCCCCACCCCAAGTCAGTGT 1948

Db        955 GATCTCTATGGAGCCATCAAGAAGCTGTGCTGCGTGCAGTCCCCAGTGCCCTCACAGGTC 1014

Qy       1949 GTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATGGCCATTGCCACCAAGATCGCC 2008

Db       1015 ATCAATGTCCGAACCATTGGTCAGCCCACCAGGCTTCGGAGCGTGGCTCAGAAAATTTTA 1074

Qy       2009 CTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGGGTGGACATGGCCCTGAAACTG 2068

Db       1075 CTTCAGATGAACTGTAAACTGGGTGGTGAGCTCTGGGGAGTGGATATTCCGCTGAAACAA 1134

Qy       2069 GCAATGATCGTGGGCATCGACTGTTACCATGACACCACAGCTGGGCGGAGGTCCATCGCA 2128

Db       1135 CTAATGGTGATTGGAATGGATGTGTACCATGACCCCAGCAGAGGCATGCGCTCTGTGGTC 1194

Qy       2129 GGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCCCGCTGCGTCTTTCAG 2188

Db       1195 GGCTTCGTGGCCAGCATAAATCTCACACTCACCAAATGGTACTCGAGGGTGGTGTTCCAG 1254

Qy       2189 GACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGCTTGCAAGCTGCTCTGAGGGCT 2248

Db       1255 ATGCCACATCAGGAGATTGTGGACAGCCTGAAGCTCTGCCTGGTGGGTTCCTTGAAAAAG 1314

Qy       2249 TGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATCGTGTACCGAGACGGTGTGGGG 2308

Db       1315 TATTATGAGGTGAACCATTGTCTCCCAGAGAAAATTGTGGTGTACCGAGATGGAGTGTCT 1374

Qy       2309 GACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCACAGTTCCTAGATTGCCTCAAG 2368

Db       1375 GATGGCCAGCTAAAGACAGTTGCCAACTACGAGATCCCTCAGCTGCAGAAGTGTTTTGAA 1434

Qy       2369 TCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATCGTGGTGAAGAAGCGTGTCAAT 2428

Db       1435 GCCTTTGATA---ACTACCACCCCAAGATGGTGGTGTTTGTAGTTCAGAAGAAAATCAGC 1491

Qy       2429 GCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAACCCTCTTCCAGGGACAGTCATC 2488

Db       1492 ACCAATCTGTACCTTGCTGCTCCTGATCACTTCGTAACCCCCTCCCCCGGGACTGTGGTT 1551

Qy       2489 GATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTCATCGTGAGCCAGGCAGTGAGA 2548

Db       1552 GATCATACCATAACCAGCTGTGAGTGGGTGGATTTCTACCTTCTTGCCCATCATGTGCGA 1611

Qy       2549 AGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTATGACAGCAGTGGCCTGAAGCCC 2608

Db       1612 CAGGGCTGTGGCATACCTACACACTACATCTGTGTTCTGAACACTGCAAATCTGAGCCCT 1671

Qy       2609 GACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTGTACTATAATTGGCCTGGAGTC 2668

Db       1672 GATCACATGCAGAGGTTGACTTTCAAACTATGCCACATGTACTGGAATTGGCCTGGTACC 1731

Qy       2669 ATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTGGCCTTCCTCGTGGGCCAGAGC 2728

Db       1732 ATCCGAGTTCCAGCTCCTTGCAAGTATGCCCACAAGCTAGCTTTCCTGTCCGGACAGATT 1791

Qy       2729 ATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTCTACTACCTCTAACCAGCAGCG 2788

Db       1792 TTGCATCATGAGCCAGCCATCCAGCTGTGTGGGAACCTGTTCTTCCTGTAACTGGGAACT 1851

Qy       2789 AGGGC 2793

Db       1852 TGGAC 1856
```

RESULT 7
US-10-102-806-267
; Sequence 267, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103P1C1
; CURRENT APPLICATION NUMBER: US/10/102,806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 267
; LENGTH: 423
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-102-806-267

Query Match 7.7%; Score 311.8; DB 15; Length 423;
Best Local Similarity 86.1%; Pred. No. 1.5e-48;
Matches 354; Conservative 2; Mismatches 54; Indels 1; Gaps 1;

```
Qy       1925 GACTGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTC 1984

Db         14 GATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTC 73

Qy       1985 ATGGCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGG 2044

Db         74 ATGGCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGRGAGCTCTGG 133

Qy       2045 CGGGTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACACC 2104

Db        134 AGGGTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTKACCATGACATG 193

Qy       2105 ACAGCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGC 2164

Db        194 ACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGC 253

Qy       2165 TGGTTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTG 2224

Db        254 TGGTTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTC 313

Qy       2225 TGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTC 2284

Db        314 TGCCTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATC 373
```

```
Qy       2285 ATCGTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAAT 2335
Db        374 ATCGTGTACCGSG-TGGCGTAGGAGACGGCCAGYTGAAAACACTGGTGACT 423
```

RESULT 8
US-09-873-737A-1
; Sequence 1, Application US/09873737A
; Patent No. US20020076797A1
; GENERAL INFORMATION:
; APPLICANT: Duke University
; APPLICANT: Lin, Haifan
; TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
; TITLE OF INVENTION: PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
; FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2
; CURRENT APPLICATION NUMBER: US/09/873,737A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/US99/28764
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/110,901
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3047
; TYPE: DNA
; ORGANISM: Drosophila sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (84)..(2612)
; NAME/KEY: misc_feature
; LOCATION: (120)
; OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (399)
; OTHER INFORMATION: n=a or t, Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (2436)
; OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
US-09-873-737A-1

Query Match 7.7%; Score 311.4; DB 10; Length 3047;
Best Local Similarity 48.7%; Pred. No. 3.8e-48;
Matches 1115; Conservative 0; Mismatches 1146; Indels 29; Gaps 9;

```
Qy        500 TCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCACTTCCGGCTGACC 559
Db        345 TCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGANTAAAA 404

Qy        560 TCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCCCTGATGGAGGCC 619
Db        405 ACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAGAAT 464

Qy        620 CGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGAAGGTGTCATGCT 679
Db        465 CCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTATCTA 524

Qy        680 TTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGTCACAGAAGTATTC 739
Db        525 TTCGACGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTCAGC 584

Qy        740 AGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCACCCTGACCAACGAGCTGCCG 799
Db        585 GGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATATCG 644

Qy        800 CCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTTCAGGAGGCTCTTGAAAATC 859
Db        645 TGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCGATGAAGGGC 704

Qy        860 ATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAGTGACCCGATTGATATTCCA 919
Db        705 CTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATAAGG 764

Qy        920 AACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCATCCTTCAGTATGAGAACAAC 979
Db        765 GAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAAGAT 824

Qy        980 ATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAGCGAGACTGTCCTAGACTTC 1039
Db        825 ATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGACATA 884

Qy       1040 ATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCAGGAGCAAGTGTCGAAGGAG 1099
Db        885 ATGCGACGTTGCTCACACAATCCGGCTCGTCA------TCAGGACGAAGTACGGGTAAAT 938

Qy       1100 CTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGACCTACCGGGTGGATGACATT 1159
Db        939 GTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGATGTC 998

Qy       1160 GACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGATGGCTCGGAGGTCAGCTTC 1219
Db        999 GACTTTGGACAAACTCCGAAATCAACATTCAGTTGCAAGGGTAG---AGATATCAGTTTC 1055

Qy       1220 CTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGACCTGAAGCAGCCGGTGCTG 1279
Db       1056 GTGGAATACTATCTCACTAAATATAATATACGCATTCGCGACCACAATCAGCCGCTGCTG 1115

Qy       1280 GTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCTGCCTGGCCCAGCTATGNTC 1339
Db       1116 ATTTCCAAA---AATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTACTA 1172

Qy       1340 ATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAATGCGCAATGATTTCAATGTG 1399
Db       1173 ATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCAGCTT 1232

Qy       1400 ATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCAGCGGCAGCGGGAGGTGGGC 1459
Db       1233 ATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATCGATTGCGC 1292

Qy       1460 CGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAGAGAGCTTCGAGACTGGGGC 1519
Db       1293 GCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGACTGGAAC 1352

Qy       1520 CTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAATCTTACAATCTGAGAAGATC 1579
Db       1353 ATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCAGAACATC 1412

Qy       1580 ----CACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGCAGACTGGTCCAAAGA 1635
Db       1413 GTGTTTCATAATGGAAAGGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTCAGA 1472

Qy       1636 AACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGGCTGCTGATCTATAC 1695
Db       1473 GACCAAAGGATGCTTACCACTCCGAGCGATGGC-CTCGATCGTTGGGCTGTCATCGCGCC 1531

Qy       1696 CCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTTCAAAGTGACTCCAGC 1755
Db       1532 GCAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAGTGG 1591

Qy       1756 CATGGGCATCCAGATGAAAAAGGCAATCA---TGATCGAGGTGGATGACAGAACAGAAGC 1812
Db       1592 AATGGGTCTTAGAATTCGAAGCCCCCAGGAATTCATAATTTATGATGATCGCACTGGAAC 1651

Qy       1813 TTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGTCTCTT 1872
Db       1652 TTATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGCCTCGT 1711

Qy       1873 GTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTGTACAGACTGCCC 1932
Db       1712 ACCCAATGATAACGCCGAAAGATACTCATCAATCAAAAAGAGAGGATACGTTGACAGGGC 1771

Qy       1933 CACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATGGCCAT 1992
Db       1772 GGTGCCAACTCA---AGTTGTGACCCTTAAAACGACCAAGAACCGTAGCCTTATGAGCAT 1828
```

```
Qy       1993 TGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGGGTGGA 2052
Db       1829 TGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGATCGA 1888

Qy       2053 CATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACACCACAGCTGG 2112
Db       1889 ACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGATCG 1948

Qy       2113 GCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTC 2172
Db       1949 GAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTACTT 2008

Qy       2173 CCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGCTTGCA 2232
Db       2009 CAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGATGAT 2068

Qy       2233 AGCT---GCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATCGT 2289
Db       2069 AGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATT 2128

Qy       2290 GTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCACA 2349
Db       2129 TTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAAGGA 2188

Qy       2350 GTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATCGT 2409
Db       2189 CATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATTAGC 2248

Qy       2410 GGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAACCC 2469
Db       2249 TTATATTGTGGTAACCAGATCCATGAA---CACGCGCTTCTTCCTCAACGGACAAAATCC 2305

Qy       2470 TCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTCAT 2529
Db       2306 TCCGCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTATCT 2365

Qy       2530 CGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTATGA 2589
Db       2366 GGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTATAG 2425

Qy       2590 CAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTGTA 2649
Db       2426 CAGCATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTTGTA 2485

Qy       2650 CTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTGGC 2709
Db       2486 CTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCTAGC 2545

Qy       2710 CTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTCTA 2769
Db       2546 TACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTTTTA 2605

Qy       2770 CTACCTCTAA 2779
Db       2606 TTATCTATAA 2615
```

RESULT 9
US-10-029-495-1
; Sequence 1, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
; APPLICANT: Drmanac, Radoje T.
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Dickson, Mark C.
; APPLICANT: Jones, Lee William
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Xue, Aidong
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi,Vinod
; TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
; FILE REFERENCE: 28110/35905A
; CURRENT APPLICATION NUMBER: US/10/029,495
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 09/378,667
; PRIOR FILING DATE: 1999-08-20
; PRIOR APPLICATION NUMBER: 09/687,527
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(559)
; OTHER INFORMATION: hiwi-Hy/CG 389
US-10-029-495-1

Query Match 4.5%; Score 183.4; DB 15; Length 560;
Best Local Similarity 58.2%; Pred. No. 1.1e-24;
Matches 316; Conservative 3; Mismatches 224; Indels 0; Gaps 0;

```
Qy       1900 TGCCATCAAGAAGTACTTGTGTACAGACTGCCCCACCCCAAGTCAGTGTGTGGTGGCCCG 1959
Db          1 TTCCATTAAAAAATATTTGAGCTYAKACTGCCCAGTCCCAAGCCAATGTGTGSTTGCTCG 60

Qy       1960 GACCCTGGGCAAGCAGCAAACAGTCATGGCCATTGCCACCAAGATCGCCCTGCAGATGAA 2019
Db         61 GACCTTGAATAAACAGGGCATGATGATGAGTATCGCCACCAAGATCGCTATGCAGATGAC 120

Qy       2020 CTGCAAGATGGGAGGCGAGCTCTGGCGGGTGGACATGGCCCTGAAACTGGCAATGATCGT 2079
Db        121 TTGCAAGCTCGGAGGCGAGCTGTGGGCTGTGGAAATACCTTTAAAGTCCCTGATGGTGGT 180

Qy       2080 GGGCATCGACTGTTACCATGACACCACAGCTGGGCGGAGGTCCATCGCAGGATTCGTCGC 2139
Db        181 CGGTATTGATGTCTGTAAAGATGCACTCAGCAAGGACGTGATGGTTGTTGGATGCGTGGC 240

Qy       2140 CAGCATCAATGAAGGGATGACCCGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCGGGCA 2199
Db        241 CAGTGTTAACCCCAGAATCACCAGGTGGTTTTCCCGCTGTATCCTTCAGAGAACAATGAC 300

Qy       2200 GGAGCTGGTGGATGGTCTCAAGGTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTG 2259
Db        301 TGATGTTGCAGATTGCTTGAAAGTTTTCATGACTGGAGCACTCAACAAATGGTACAAGTA 360

Qy       2260 CAATGAATACATGCCCAGCCGTGTCATCGTGTACCGAGACGGTGTGGGGGACGGGCAGCT 2319
Db        361 CAATCATGATTTGCCAGCACGGATAATTGTGTACCGTGCTGGTGTAGGGGATGGTCAGCT 420

Qy       2320 GAAGACCCTGGTCAATTATGAGGTCCCACAGTTCCTAGATTGCCTCAAGTCAGTCGGGAG 2379
Db        421 GAAAACACTTATTGAATATGAAGTCCCACAGCTGCTGAGCAGTGTGGCAGAATCCAGCTC 480

Qy       2380 AGGTTACAACCCAAGACTGACTGTAATCGTGGTGAAGAAGCGTGTCAATGCCAGGTTTTT 2439
Db        481 AAATACCAGCTCAAGACTGTCGGTGATTGTGGTCAGGAAGAAGTGCATGCCACGATTCTT 540

Qy       2440 TGC 2442
Db        541 TAC 543
```

RESULT 10
US-10-027-632-166243
; Sequence 166243, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632

```
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 166243
;   LENGTH: 889
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-166243

  Query Match             3.6%;  Score 144.8;  DB 15;  Length 889;
  Best Local Similarity   84.7%;  Pred. No. 1.9e-17;
  Matches  161;  Conservative    1;  Mismatches   28;  Indels    0;  Gaps    0;

Qy       1597 GTTTGATTACAACCCACAATTTGCAGACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGAT 1656
             ||||||||||| |||||||||||||| |||||||||||||||||||| || ||  | ||
Db        215 GTTTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAACAAGAGGTGCACCATTAAT 274

Qy       1657 CAGCGTGAAGCCATTGGATAACTGGCTGCTGATCTATACCCGCAGGAATTATGAAGCAGC 1716
               || || |||||| | |||||||||||| |||||||||| || || ||||||||||||||
Db        275 TAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACRCGAAGAAATTATGAAGCAGC 334

Qy       1717 CAACTCACTGATACAGAACCTGTTCAAAGTGACTCCAGCCATGGGCATCCAGATGAAAAA 1776
             ||| ||| ||||||| || || || ||||| || |||||||||||||| || ||||:|||
Db        335 CAATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGCATGCAAATGARAAA 394

Qy       1777 GGCAATCATG 1786
              ||||| |||
Db        395 AGCAATAATG 404


RESULT 11
US-10-027-632-166244
; Sequence 166244, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 166244
;   LENGTH: 889
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-166244

  Query Match             3.6%;  Score 144.8;  DB 15;  Length 889;
  Best Local Similarity   84.7%;  Pred. No. 1.9e-17;
  Matches  161;  Conservative    1;  Mismatches   28;  Indels    0;  Gaps    0;

Qy       1597 GTTTGATTACAACCCACAATTTGCAGACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGAT 1656
             ||||||||||| |||||||||||||| |||||||||||||||||||| || ||  | ||
Db        215 GTTTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAACAAGAGGTGCACCATTAAT 274

Qy       1657 CAGCGTGAAGCCATTGGATAACTGGCTGCTGATCTATACCCGCAGGAATTATGAAGCAGC 1716
               || || |||||| | |||||||||||| |||||||||| || || ||||||||||||||
Db        275 TAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACRCGAAGAAATTATGAAGCAGC 334

Qy       1717 CAACTCACTGATACAGAACCTGTTCAAAGTGACTCCAGCCATGGGCATCCAGATGAAAAA 1776
             ||| ||| ||||||| || || || ||||| || |||||||||||||| || ||||:|||
Db        335 CAATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGCATGCAAATGARAAA 394

Qy       1777 GGCAATCATG 1786
              ||||| |||
Db        395 AGCAATAATG 404


RESULT 12
US-10-027-632-166245
; Sequence 166245, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 166245
;   LENGTH: 889
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-166245

  Query Match             3.6%;  Score 144.8;  DB 15;  Length 889;
  Best Local Similarity   84.7%;  Pred. No. 1.9e-17;
  Matches  161;  Conservative    1;  Mismatches   28;  Indels    0;  Gaps    0;

Qy       1597 GTTTGATTACAACCCACAATTTGCAGACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGAT 1656
             ||||||||||| |||||||||||||| |||||||||||||||||||| || ||  | ||
Db        215 GTTTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAACAAGAGGTGCACCATTAAT 274

Qy       1657 CAGCGTGAAGCCATTGGATAACTGGCTGCTGATCTATACCCGCAGGAATTATGAAGCAGC 1716
               || || |||||| | |||||||||||| |||||||||| || || ||||||||||||||
Db        275 TAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACRCGAAGAAATTATGAAGCAGC 334

Qy       1717 CAACTCACTGATACAGAACCTGTTCAAAGTGACTCCAGCCATGGGCATCCAGATGAAAAA 1776
             ||| ||| ||||||| || || || ||||| || |||||||||||||| || ||||:|||
Db        335 CAATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGCATGCAAATGARAAA 394

Qy       1777 GGCAATCATG 1786
```

```
                  ||||| |||
Db        395 AGCAATAATG 404


RESULT 13
US-09-960-352-1281
; Sequence 1281, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 1281
;   LENGTH: 444
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 06-LIB3058-048-Q1-K1-B9
US-09-960-352-1281

  Query Match             3.1%;  Score 125;  DB 11;  Length 444;
  Best Local Similarity   72.9%;  Pred. No. 6.9e-14;
  Matches  161;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy       3844 TTTGACAAAGAGAGCAGCTTAGAGTGGCTAGAGTGTGATGAGTTTTACACAATGGCGATG 3903
              | |  ||| ||  |  |     ||  ||  || | |  ||   |||  |||  ||| |
Db         40 TCTCTCAAGGAAGGTGGGGACAAGAAGCAGGATTCTTTTGTACTTTGTACATAGGCTACA 99

Qy       3904 TGTGTGTGTGCCCCTGGCACAGCTCAATAAAGTTTCTGTCTCACTTTAAAAAAAAAAAAA 3963
                | |||||   |          | ||  |  ||| | | ||     ||||||||||||
Db        100 CATTTGTGTAAACAGTTTTTGAATAAAATATTTTTTTTTTTCGTAAAAAAAAAAAAAAAA 159

Qy       3964 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4023
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        160 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 219

Qy       4024 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4064
              ||||||||||||||||||||||||||||||||||||||||
Db        220 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 260


RESULT 14
US-09-908-711-53
; Sequence 53, Application US/09908711
; Patent No. US20020045230A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA128
;  CURRENT APPLICATION NUMBER: US/09/908,711
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US01/01360
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 09/764,867
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: US01/01344
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 09/764,892
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: US01/01345
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 09/764,888
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: US01/01329
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 09/764,905
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: US01/01354
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 09/764,891
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: US01/01339
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 09/764,869
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: US01/01340
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 09/764,874
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: US01/01334
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 09/764,898
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: US01/01320
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 09/764,853
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: US01/01349
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 09/764,902
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: US01/01239
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 09/764,870
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: US01/01348
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 09/764,882
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: US01/01347
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 09/764,896
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: US01/01307
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 09/764,864
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: US01/01341
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 09/764,856
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: US01/01336
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 09/764,868
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: US01/01312
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 60/179,065
;  PRIOR FILING DATE: 2000-01-31
;  PRIOR APPLICATION NUMBER: 60/180,628
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: 60/209,467
;  PRIOR FILING DATE: 2000-06-07
;  NUMBER OF SEQ ID NOS: 167
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
;   LENGTH: 1225
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-908-711-53

  Query Match             3.0%;  Score 123.8;  DB 10;  Length 1225;
  Best Local Similarity   94.8%;  Pred. No. 1.7e-13;
  Matches  128;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       3928 CAATAAAGTTTCTGTCTCACTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3987
              ||||||||||| |  ||| | |  ||||||||||||||||||||||||||||||||||||
Db       1091 CAATAAAGTTTATTCCTCCCCTGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1150

Qy       3988 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4047
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1210

Qy       4048 AAAAAAAAAAAAAAA 4062
                 |||||||||||||||
Db       1211 AAAAAAAAAAAAAAA 1225


RESULT 15
US-10-106-698-464
; Sequence 464, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
;  APPLICANT: Ruben et al.
;  TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
;  FILE REFERENCE: PA005P1
;  CURRENT APPLICATION NUMBER: US/10/106,698
;  CURRENT FILING DATE:  2002-03-27
;  PRIOR APPLICATION NUMBER: PCT/US00/26524
;  PRIOR FILING DATE: 2000-09-28
;  PRIOR APPLICATION NUMBER: US 60/157,137
;  PRIOR FILING DATE: 1999-09-29
;  PRIOR APPLICATION NUMBER: US 60/163,280
;  PRIOR FILING DATE: 1999-11-03
;  NUMBER OF SEQ ID NOS: 8564
;  SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 464
;   LENGTH: 2288
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-106-698-464

  Query Match             3.0%;  Score 123.8;  DB 15;  Length 2288;
  Best Local Similarity   94.8%;  Pred. No. 2.2e-13;
  Matches  128;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       3930 ATAAAGTTTCTGTCTCACTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3989
                 ||||||||    | |||||  ||||||||||||||||||||||||||||||||||||||||||
Db       2142 ATAAAGTTGTATTTTCACTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2201

Qy       3990 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4049
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2202 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2261

Qy       4050 AAAAAAAAAAAAAAA 4064
                 |||||||||||||||
Db       2262 AAAAAAAAAAAAAAA 2276


Search completed: July 22, 2003, 06:59:25
Job time : 787.537 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2003, 13:46:42 ; Search time 5337 Seconds
(without alignments)
12332.482 Million cell updates/sec

Title: US-09-873-737A-3
Perfect score: 4064
Sequence: 1 agaagcctgctacaggggt.....aaaaaaaaaaaaaaaaaaa 4064

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	785.2	19.3	978	9	AV253931	AV253931 AV253931
2	717.6	17.7	971	10	BB614499	BB614499 BB614499
3	555	13.7	565	12	BG101107	BG101107 uy19e03.y
4	445.4	11.0	818	12	BG773137	BG773137 602721862
5	433.2	10.7	874	12	BF666315	BF666315 602119233
6	430.4	10.6	577	12	BF080918	BF080918 233684 MA
7	426.6	10.5	779	13	BI560710	BI560710 603254744
8	391.2	9.6	533	12	BF080923	BF080923 233690 MA
9	387.2	9.5	1495	12	BF664941	BF664941 602119191
10	383.6	9.4	1092	13	BI463482	BI463482 603203560
11	370.8	9.1	422	10	BB709598	BB709598 BB709598
12	370	9.1	882	13	BI459464	BI459464 603200456
13	364.6	9.0	622	9	AL705072	AL705072 DKFZp686E
14	363.6	8.9	527	14	BM751055	BM751055 K-EST0026
15	363.4	8.9	666	13	BI463897	BI463897 603202635
16	358.8	8.8	640	9	AL705301	AL705301 DKFZp686G
17	358.4	8.8	715	12	BG721550	BG721550 602695155
18	356	8.8	705	12	BF246708	BF246708 601855507
19	346	8.5	731	13	BI458664	BI458664 603199792
20	323	7.9	462	9	AA969938	AA969938 op60a11.s
21	314.4	7.7	581	12	BG718168	BG718168 602696127
c 22	304.8	7.5	413	9	AL041473	AL041473 DKFZp434E
23	295.6	7.3	605	9	AL704986	AL704986 DKFZp686O
24	289.8	7.1	477	13	BM539361	BM539361 hb08e02.g
25	286.8	7.1	609	14	BQ481304	BQ481304 faa74e09.
26	286.4	7.0	539	13	BI463615	BI463615 603206270
27	277.8	6.8	420	14	BQ374910	BQ374910 MR4-TN010
28	276.8	6.8	587	9	AA430311	AA430311 zw68h01.r
29	276.6	6.8	744	13	BI459142	BI459142 603199321
30	267.8	6.6	568	14	BQ074332	BQ074332 fz27h06.y
31	266.4	6.6	588	14	BQ449946	BQ449946 faa63a07.
32	262	6.4	534	13	BI672860	BI672860 ft38d09.y
33	253.2	6.2	935	12	BF247005	BF247005 601855750
34	241.6	5.9	249	9	AV274252	AV274252 AV274252
35	240.4	5.9	612	13	BM141396	BM141396 ZFT260 Ze
36	238.4	5.9	617	13	BJ040023	BJ040023 BJ040023
37	235.4	5.8	795	12	BG191694	BG191694 RST10790
38	234.6	5.8	474	9	AL705611	AL705611 DKFZp686L
39	232.8	5.7	588	10	AV668679	AV668679 AV668679
40	229.6	5.6	785	12	BG221179	BG221179 RST40982
41	229	5.6	800	12	BG206124	BG206124 RST25559
42	228	5.6	796	12	BG209773	BG209773 RST29299
43	227.8	5.6	239	9	AV276025	AV276025 AV276025
44	226	5.6	778	12	BG210801	BG210801 RST30349
45	226	5.6	783	12	BG198827	BG198827 RST18097

ALIGNMENTS

RESULT 1
AV253931
LOCUS AV253931 978 bp mRNA linear EST 24-OCT-2001
DEFINITION AV253931 RIKEN full-length enriched, adult male testis (DH10B) Mus musculus cDNA clone 4921506H11 3', mRNA sequence.
ACCESSION AV253931
VERSION AV253931.2 GI:16387973
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 978)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT On Nov 4, 1999 this sequence version replaced gi:6241390.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222

Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES Location/Qualifiers
source 1. .978
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4921506H11"
/clone_lib="RIKEN full-length enriched, adult male testis (DH10B)"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."

BASE COUNT 217 a 221 c 237 g 303 t
ORIGIN

Query Match 19.3%; Score 785.2; DB 9; Length 978;
Best Local Similarity 94.3%; Pred. No. 2.9e-90;
Matches 862; Conservative 0; Mismatches 38; Indels 14; Gaps 4;

```
Qy       2729 ATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTCTACTACCTCTAACCAGCAGCG 2788
             | |  || |||| ||||||||||    | | ||  ||   ||||||||||||||||||
Db         79 AGCATCACAGAGACAAACCTCTCCTGTCCAACCCCTCT---CTACCTCTAACCAGCAGCG 135

Qy       2789 AGGGCCGGCATGCTGGTTTCCTTCAGAGCTCTGCCCAATCCTTCAGTTCTATTTTTCATA 2848
             | |||||  ||||||  | |||  ||    ||   | |  | ||||| ||||||||||||
Db         136 A-GGCCGCGATGCTGTCTCCCTCAAG----CTCTGCCAATCCTCAGTACTATTTTTCATA 190

Qy       2849 GGATGAGGTTCTAGCGGGATGTTGGGGGTAGGGGTTGGGGGCCGCTGTTTTGTTTTGTTT 2908
             |||||  | |||||||  |||||  ||||||||||| | ||||||||||||||||||||
Db         191 CGATGATGCTCTAGCGCGATGTGTGGGGTAGGGGTGGAGGGCCGCTGTTTTGTTTTGTTT 250

Qy       2909 TGTTTTTAGTTGTTGGTGTGTAGCGGCTATAGAGTTAACTAAGGCTTATGTTTGCTTTGT 2968
             |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db         251 TGTTTTTAGTTGTTGGTGTGTAGCGGCTATAGAGATAACTAAGGCTTATGTTTGCTTTGT 310

Qy       2969 TTCTCATGGTGTGTTGTGACAGTTTGCTTGGCAGAGTGGTGGCAAGGGCCGTGGTACCAC 3028
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         311 CTCTCATGGTGTGTTGTGACAGTTTGCTTGGCAGAGTGGTGGCAAGGGCCGTGGTACCAC 370

Qy       3029 AGATAGACTTTCCTAAAGCAGCCGTGTTGGCAGGTTCTTGGAGGTGCTCCTGGGAGCACC 3088
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         371 AGATAGACTTTCCTAAAGCAGCCGTGTTGGCAGGTTCTTGGAGGTGCTCCTGGGAGCACC 430

Qy       3089 TCCTATTGTACGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 3148
             |||||||||||||||      ||||||||||||||||||||||||||||||||||||||||
Db         431 TCCTATTGTACGCTG------CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 484

Qy       3149 TCTCTCTCTGCATGTTTGAATTTAAAAGAGGAAAGGACACAAGGCAAGAATGCCCAGAGC 3208
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         485 TCTCTCTCTGCATGTTTGAATTTAAAAGAGGAAAGGACACAAGGCAAGAATGCCCAGAGC 544

Qy       3209 ATCCTAGATGGGTGCTATATGGGACTCTGACACCTCAGGTTCTTAAAGATAAAAAGGCCA 3268
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         545 ATCCTAGATGGGTGCTATATGGGACTCTGACACCTCAGGTTCTTAAAGATAAAAAGGCCA 604

Qy       3269 GAGGTACCTTTCTGTAATTCTTTTAAGAGTTAGGTCATAAGGAATGAAAAGAGTTACTTT 3328
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         605 GAGGTACCTTTCTGTAATTCTTTTAAGAGTTAGGTCATAAGGAATGAAAAGAGTTACTTT 664

Qy       3329 TCTCCCTCCATGAAAGTGTTTTTTAAGACTGACAAGGCCGGTGGTGGAGTGAGTGTTGTT 3388
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         665 TCTCCCTCCATGAAAGTGTTTTTTAAGACTGACAAGGCCGGTGGTGGAGTGAGTGTTGTT 724

Qy       3389 GTTTTGTTTCGCAATTTTCTTACCAAGTGAAATTACTGACCATAATTTTATTTAATAGTT 3448
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         725 GTTTTGTTTCGCAATTTTCTTACCAAGTGAAATTACTGACCATAATTTTATTTAATAGTT 784

Qy       3449 TTTAGTTTTTGAAAGCCTGTGTTTTCTTTGCTTTTAAACTACAGGCGACCACTCTGGGAT 3508
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         785 TTTAGTTTTTGAAAGCCTGTGTTTTCTTTGCTTTTAAACTACAGGCGACCACTCTGGGAT 844

Qy       3509 AGGATTCCACTCCCAAAAGTAATGTGTTGGGTGTGGTGGCACACGCCTGTGATCTCAGGA 3568
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         845 AGGATTCCACTCCCAAAAGTAATGTGTTGGGTGTGGTGGCACACGCCTGTGATCTCAGGA 904

Qy       3569 CTCAGGAGGTGAAGCCAGGGGAACCCCAAGTTCCCGGTCAGCCTAGGCTACATAGTGAGA 3628
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         905 CTCAGGAGGTGAAGCCAGGGGAACCCCAAGTTCCCGGTCAGCCTAGGCTACATAGTGAGA 964

Qy       3629 CTTTGTCTTAACAA 3642
             ||||||||||||||
Db         965 CTTTGTCTTAACAA 978
```

RESULT 2
BB614499
LOCUS BB614499 971 bp mRNA linear EST 26-OCT-2001
DEFINITION BB614499 RIKEN full-length enriched, adult male testis Mus musculus cDNA clone 4921506H11 5', mRNA sequence.
ACCESSION BB614499
VERSION BB614499.1 GI:16454917
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 971)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.

, Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

JOURNAL Unpublished (2001)

COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers

source 1. .971
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4921506H11"
/clone_lib="RIKEN full-length enriched, adult male testis"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="SOLR"
/note="Site_1: XhoI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAGCGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. "

BASE COUNT 235 a 264 c 296 g 173 t 3 others

ORIGIN

Query Match 17.7%; Score 717.6; DB 10; Length 971;
Best Local Similarity 93.8%; Pred. No. 9.8e-82;
Matches 767; Conservative 0; Mismatches 48; Indels 3; Gaps 2;

```
Qy          1 AGAAGCCTGCTACAGGGGGTGGCGTGAGGAGCCTCAAGGACTGCGGACCTAAGCGTGCTA 60
              |||||||||| |||||||||||||||||||||||| |||||||||| ||||||||||||
Db         69 AGAAGCCTGCTGCAGGGGGTGGCGTGAGGAGCCTCAGGGACTGCGGANCTAAGCGTGCTA 128

Qy         61 AGCACGGGGCCAGGATAGGACTGGACGCTGCTGGACCACGACGATCAGGGAGTGACCCCC 120
              | |||||||||||||||||||||||| |||||||||||||||| ||||||||||||||
Db        129 ANCACGGGGCCAGGATAGGACTGGACCCTGCTGGACCACGACGAGTAGGGAGTGACCCCC 188

Qy        121 ACCAGCATCCAGCCTGCGGCAGTTGTAGGACTGGAAGGAGGCCAAGCTGAACACCTGAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        189 ACCAGCATCCAGCCTGCGGCAGTTGTAGGACTGGAAGGAGGCCAAGCTGAACACCTGAGC 248

Qy        181 TGACTGGAAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        249 TGACTGGAAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGA 308

Qy        241 GACGGTGCAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        309 GACGGTGCAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACC 368

Qy        301 TCAACAGTCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGT 360
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        369 TCAACGGTCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGT 428

Qy        361 CGGAGCCACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        429 CGGAGCCACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACT 488

Qy        421 GGCAGAGAGAGGAGGGCGTCGCCGAGACTTCCATGACNTTGGTGTGAACACCAGACAGAA 480
              |||||||||||||||||||| | |||||||||||  ||||||||||||||||||||||
Db        489 GGCAGAGAGAGGAGGGCGTCGTCTAGACTTCCATGAGCTTGGTGTGAACACCAGACAGAA 548

Qy        481 CCTTGACCATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCAC 540
              |||||||||||||||||||| |||||||| ||||||||||||||||||||||||||||||
Db        549 CCTTGACCATGTCAAAGAGTCACAGACAGGCGCCTCTGGCATCATTGTGAAGCTGAGCAC 608

Qy        541 CAACCACTTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTA 600
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        609 CAACCACTTCCGGGTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTA 668

Qy        601 CAATCCCCTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCT 660
              ||||||| ||||||||| || ||  |    ||||||||||||||||||||||||||||||
Db        669 CAATCCCTTGATGGAGG--CGNAGAGCTTCGCTCGCACTGCTCTTCCAGCATGAAGACCT 726

Qy        661 CATTGGAAGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCA 720
              |||||||||||||||||||||||||||||||| |||||||||||||| |||||||||
Db        727 CATTGGAAGGTGTCATGCTTTCGATGGGACAATAATGTTTTTACCTAAGATACTACAGCA 786

Qy        721 CAAGGTCACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCAC 780
              ||||||| ||||||| |||||||||||||||||||||||||||||||| || ||||||||
Db        787 CAAGGTCGAAGAAGTAGTCAGTCAGACTCGGAATGGGGAACACGTGAGCAT-ACCATCAC 845

Qy        781 CCTGACCAACGAGCTGCCGCCCACCTCGCCCACCTGCC 818
              ||| |||||    |  |  | |  | |  ||  |||||
Db        846 CCTTACCAAGAGCCGCCGCCACCTCGCCACCTGCTGCC 883
```

RESULT 3
BG101107
LOCUS BG101107 565 bp mRNA linear EST 29-JAN-2001
DEFINITION uy19e03.y1 McCarrey Eddy spermatocytes Mus musculus cDNA clone IMAGE:3658468 5' similar to TR:O95404 O95404 HIWI ;, mRNA sequence.
ACCESSION BG101107
VERSION BG101107.1 GI:12596424
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 565)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person

```
           ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
           ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
           Waterston,R. and Wilson,R.
  TITLE    The WashU-NCI Mouse EST Project 1999
  JOURNAL  Unpublished (1999)
COMMENT    Contact: Marra M/WashU-NCI Mouse EST Project 1999
           Washington University School of Medicine
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: mouseest@watson.wustl.edu
           This clone is available royalty-free through LLNL ; contact the
           IMAGE Consortium (info@image.llnl.gov) for further information.
           MGI:1420772
           Seq primer: Primer name ambiguous
           High quality sequence stop: 418.
FEATURES             Location/Qualifiers
     source          1. .565
                     /organism="Mus musculus"
                     /strain="CD-1"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3658468"
                     /clone_lib="McCarrey Eddy spermatocytes"
                     /sex="male"
                     /tissue_type="spermatocytes, pooled from multiple mice"
                     /dev_stage="60 day"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: testis; Vector: pBluescript SK+ (Stratagene
                     ); Site_1: XhoI; Site_2: EcoRI; cDNA oligo dT-primed
                     [5'-(GA)10-ACTAGTCTCGAGTTTTTTTTTTTT-3'] and directionally
                     cloned  using 5' linkers 5'-AATTCGGCACGAG-3' and
                     5'-CTCGTGCCG-3'. Size selection of >400bp material gives
                     average insert size ranging from 1-2 kb. Library was mass
                     excised (from  lambda-UniZAP-XR) and resulting
                     single-stranded phagemids were prepped and tranformed
                     into DH10B. Library contains 98% recombinants.
                     References: J. Androl. 20:635-639 and Gene 25:263-269.
                     Library constructed and donated by J. McCarrey, Ph.D.
                     (Southwest Foundation for Biomedical Research, Dept. of
                     Genetics); excision done by E.M. Eddy, Ph.D. (National
                     Institutes of  Health, National Institute of Environmental
                     Health Sciences). Original lambda-based library is
                     available through ATCC, catalog #63422."
BASE COUNT      128 a    140 c    168 g    129 t
ORIGIN

  Query Match             13.7%;  Score 555;  DB 12;  Length 565;
  Best Local Similarity   98.9%;  Pred. No. 4.4e-61;
  Matches  558;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       2128 AGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCCCGCTGCGTCTTTCA 2187
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCCCGCTGCGTCTTTCA 60

Qy       2188 GGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGCTTGCAAGCTGCTCTGAGGGC 2247
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGCTTGCAAGCTGCTCTGAGGGC 120

Qy       2248 TTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATCGTGTACCGAGACGGTGTGGG 2307
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATCGTGTACCGAGACGGTGTGGG 180

Qy       2308 GGACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCACAGTTCCTAGATTGCCTCAA 2367
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCACAGTTCCTAGATTGCCTCAA 240

Qy       2368 GTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATCGTGGTGAAGAAGCGTGTCAA 2427
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATCGTGGTGAAGAAGCGTGTCAA 300

Qy       2428 TGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAACCCTCTTCCAGGGACAGTCAT 2487
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAACCCTCTTCCAGGGACAGTCAT 360

Qy       2488 CGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTCATCGTGAGCCAGGCAGTGAG 2547
             ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db        361 CGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTGTCATCGTGAGCCAGGCAGTGAG 420

Qy       2548 AAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTATGACAGCAGTGGCCTGAAGCC 2607
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTATGACAGCAGTGGCCTGAAGCC 480

Qy       2608 CGACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTGTACTATAATTGGCCTGGAGT 2667
             ||||||||| |||||||||||||||||| |||| ||||||||||||||||||||||||||
Db        481 CGACCACATTCAGCGGCTGACATACAAGCTCTGTCACGTGTACTATAATTGGCCTGGAGT 540

Qy       2668 CATCCGAGTCCCTGCACCTTGCCA 2691
             ||| ||||| ||||||||||||||
Db        541 CATGCGAGTTCCTGCACCTTGCCA 564


RESULT 4
BG773137
LOCUS       BG773137                 818 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602721862F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4838918 5',
            mRNA sequence.
ACCESSION   BG773137
VERSION     BG773137.1  GI:14083790
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 818)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10773  row: k  column: 15
            High quality sequence stop: 789.
FEATURES             Location/Qualifiers
     source          1. .818
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4838918"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      245 a    174 c    198 g    201 t
ORIGIN

  Query Match             11.0%;  Score 445.4;  DB 12;  Length 818;
  Best Local Similarity   78.0%;  Pred. No. 2.3e-47;
  Matches  612;  Conservative    0;  Mismatches  163;  Indels   10;  Gaps    6;

Qy        188 AAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGAGACGGTG 247
              | |||||||||  ||||||||||  | ||| | || ||||| || ||||||||||| | |
Db         37 ACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCG 96
```

```
Qy        248 CAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACCTCAACAG 307
              ||||  || ||  |  |||| || |||||||||||| || ||  |  || || ||||| | |
Db         97 CAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCG 156

Qy        308 TCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGTCGGAGCC 367
              | ||  |||||||||  ||  ||||||| ||||| ||||| || |    ||  ||||
Db        157 CCACCAGCAGAGGGGGGATTATTTGGCCGTGGACGGCAGAGAGGAA---CAGCAGGAGGA 213

Qy        368 ACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACTGGCAGAG 427
              ||| |||||||||||| ||| ||||| || ||||| ||||||||| | ||  | ||||||
Db        214 ACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTAGCAGAG 273

Qy        428 AGAGGAGGGCGTCGCCGAGACTTCCATGACNTTGGTGTGAACACCAGACAGAACCTTGAC 487
              |||||||| |||||  |||| || |||||  |||||||||| || || |||||||| |||
Db        274 AGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGAC 333

Qy        488 CATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCAC 547
              ||||| ||||| ||||| ||||| || || ||||| || || | | | ||||| |||||
Db        334 CATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCAT 393

Qy        548 TTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCC 607
              |||||||||| || || || ||||||||| | |||||||| ||||| ||||| || ||
Db        394 TTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCA 453

Qy        608 CTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGA 667
              |||||||| ||| |||| || || || || || || || || || ||||| || ||||||
Db        454 CTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGA 513

Qy        668 AGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGTC 727
              | |||||||||||| ||||| || ||||| |||||||||| |||||||||| |||||
Db        514 AAGTGTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTT 573

Qy        728 ACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCACCCTGACC 787
              || ||||| || ||| |||| |||||||| ||  | |||||||| || |||||  | ||
Db        574 ACTGAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACA 633

Qy        788 AA-CGAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTTCAGGAG 846
              ||  || || || || || || || || ||  |||||||||||| ||  | ||||||||
Db        634 AATTGAACTTCCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAG 693

Qy        847 GC-TCTTGAAAATCATGAATTTGC---AACAAATTGGACGGAATTATTACAATCCAAGTG 902
              || | ||||||||||||||||||    | || | ||||| |||||||| || |||| ||
Db        694 GCTTTTTGAAAATCATGAATTTGCCAACAAAATTGGGACGAAATTATTATAACCCAAATG 753

Qy        903 ACCCGA-TTGATATTCCAAACCACAGGTTGGTGATCTGGCCC-GGCTTCACCACCTCCAT 960
              |||| | ||||||||||||  |||||||||||||| |||||| |||||||| |  |||||
Db        754 ACCCAATTTGATATTCCAAGTCACAGGTTGGTGATTTGGCCCTGGCTTCACTATTTCCAT 813

Qy        961 CCTTC 965
              |||||
Db        814 CCTTC 818
```

RESULT 5
BF666315

```
LOCUS       BF666315               874 bp    mRNA    linear   EST 21-DEC-2000
DEFINITION  602119233F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4276136 5',
            mRNA sequence.
ACCESSION   BF666315
VERSION     BF666315.1  GI:11940210
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 874)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1096  row: b  column: 09
            High quality sequence stop: 594.
FEATURES             Location/Qualifiers
     source          1. .874
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4276136"
                     /clone_lib="NIH_MGC_56"
                     /tissue_type="primitive neuroectoderm"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA.  5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.65 kb (range 0.9-4.0 kb).  15/15 colonies
                     contained  inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      200 a    218 c    258 g    198 t
ORIGIN
```

```
  Query Match             10.7%;  Score 433.2;  DB 12;  Length 874;
  Best Local Similarity   82.8%;  Pred. No. 7.7e-46;
  Matches  543;  Conservative    0;  Mismatches  108;  Indels    5;  Gaps    4;

Qy       1784 ATGATCGAGGTGGATGACAGAACAGAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACG 1843
              | ||| || |||||||||||||| ||||| ||  | |||| ||| ||||| ||||| ||
Db          8 AGGATTGAAGTGGATGACAGAACTGAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACA 67

Qy       1844 TCAGACACTCAGATAGTTGTCTGTCTCTTGTCAAGTAATCGGAAGGACAAATATGATGCC 1903
              ||||||| |||||||||||||||| ||||||||||||||||||||||||||||||| |||||
Db         68 GCAGACACCCAGATAGTTGTCTGTCTGTTGTCAAGTAATCGGAAGGACAAATACGATGCT 127

Qy       1904 ATCAAGAAGTACTTGTGTACAGACTGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACC 1963
              || || || ||| |||||||||| ||||| |||||||||||||||||||||||||||| |||
Db        128 ATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACC 187

Qy       1964 CTGGGCAAGCAGCAAACAGTCATGGCCATTGCCACCAAGATCGCCCTGCAGATGAACTGC 2023
              | ||||| |||||||| |||||||||||||| || ||||| ||||| ||||||||||||
Db        188 TTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAAAGATTGCCCTACAGATGAACTGC 247

Qy       2024 AAGATGGGAGGCGAGCTCTGGCGGGTGGACATGGCCCTGAAACTGGCAATGATCGTGGGC 2083
              |||||||||| |||||||| |||||||||  ||||||| || |  |||||||| |||
Db        248 AAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGC 307

Qy       2084 ATCGACTGTTACCATGACACCACAGCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGC 2143
              ||||| ||||||||||||||  |||||||||||||||||| |||||||||| || ||||||
Db        308 ATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGC 367

Qy       2144 ATCAATGAAGGGATGACCCGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAG 2203
              |||||||||||||||||||||||||||||||||| |||||| | ||||||||  | || ||||||
Db        368 ATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCATATTTCAGGATAGAGGACAGGAG 427

Qy       2204 CTGGTGGATGGTCTCAAGGTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAAT 2263
              ||||| ||||| ||||| || ||| ||||||| | |||||||||||||| | ||||||||
Db        428 CTGGTAGATGGGCTCAAAGTCTGCCTGCAAGC-GGTCTGAGGGCTTGGAATAGCTGCAAT 486

Qy       2264 GAATACATGCCCAGCCGTGTCATCGTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAG 2323
              || |||||||||||||  |||||||||||| || || || || |||||     ||||
Db        487 GAGTACATGCCCAGCCGGATCATCGTGTACCGCGATGGCGTAGGAGACGG--CAGTGAAA 544

Qy       2324 ACCCTGGTCAATTATGAGGTCCCACAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGT 2383
              || ||||| || || || || |||||||  | ||| | || || || |  | ||||||
Db        545 ACACTGGTGAACTACGAAGTGCCACAGTT-TTGGATAGTCTAAAATCCATGGTTAGAGGT 603
```

```
Qy       2384 TACAACCCAAGACTGACTGTAATCGTGGTGAAGAAGCGTGTCAATGCCAGGTTTTT 2439
             | |||||| ||||| || ||| |||||||| |  | || ||  |  ||  | ||||
Db        604 T-CAACCCTAGACTAACGGTACTCGTGGTGGACGACCGAGTGGACCCCCAGATTTT 658


RESULT 6
BF080918
LOCUS       BF080918                577 bp    mRNA    linear   EST 18-OCT-2000
DEFINITION  233684 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF080918
VERSION     BF080918.1  GI:10874739
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 577)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 53  row: N  column: 21
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .577
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      198 a     98 c    132 g    149 t
ORIGIN

  Query Match           10.6%;  Score 430.4;  DB 12;  Length 577;
  Best Local Similarity   84.2%;  Pred. No. 2.3e-45;
  Matches  485;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy       1583 CAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGCAGACTGGTCCAAAGAAACAAGA 1642
             || || ||||| || |||||||||| |||||||||||||| ||||| ||||||||||||
Db          2 CAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCAAAAGAAACAAGA 61

Qy       1643 GGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGGCTGCTGATCTATACCCGCAGG 1702
             || || ||  | || || || |||||| | |||||||||||| |||||||||| || ||
Db         62 GGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACTCGAAGA 121

Qy       1703 AATTATGAAGCAGCCAACTCACTGATACAGAACCTGTTCAAAGTGACTCCAGCCATGGGC 1762
             |||||||||||||||| ||| ||||||| || || || ||||| || ||||||||||||
Db        122 AATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGC 181

Qy       1763 ATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGACAGAACAGAAGCTTATCTGAGA 1822
             || || |||||||| ||||| ||||| || |||||||| ||||| ||||| ||| | |||
Db        182 ATACAAATGAAAAAAGCAATAATGATTGAAGTGGATGATAGAACTGAAGCCTATTTAAGA 241

Qy       1823 GCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGTCTCTTGTCAAGTAAT 1882
             | ||| ||||| ||||| || || || || |||||||||| ||||| ||||||||||||
Db        242 GTCTTACAGCAAAAGGTTACATCGGATACCCAGATAGTTGTTTGTCTGTTGTCAAGTAAT 301

Qy       1883 CGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTGTACAGACTGCCCCACCCCAAGT 1942
             || |||||||||||||||| || || || ||||| |||||||| ||||| || ||||||
Db        302 CGAAAGGACAAATATGATGCTATTAAAAAATACTTATGTACAGATTGCCCTACTCCAAGT 361

Qy       1943 CAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATGGCCATTGCCACCAAG 2002
             |||||||||||||||| ||| | ||||| || || || ||||||||||||||| || |||
Db        362 CAGTGTGTGGTGGCCCGAACCTTAGGCAAACAACAGACTGTCATGGCCATTGCGACAAAG 421

Qy       2003 ATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGGGTGGACATGGCCCTG 2062
             |||||||||||||||||||||||||||||||| ||||| ||| |||| || ||| | ||
Db        422 ATCGCCCTGCAGATGAACTGCAAGATGGGAGGAGAGCTTTGGAGGGTTGATATGCCTCTA 481

Qy       2063 AAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACACCACAGCTGGGCGGAGGTCC 2122
             || || || ||||| || ||||| || ||||| ||||| |||||||||||| ||||||||
Db        482 AAGCTAGCGATGATAGTTGGCATTGATTGTTATCATGATACCACAGCTGGACGGAGGTCA 541

Qy       2123 ATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATG 2158
             || |||||||| || || ||||||||||||||||||
Db        542 ATTGCAGGATTTGTTGCAAGCATCAATGAAGGGATG 577


RESULT 7
BI560710
LOCUS       BI560710                779 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  603254744F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5296954 5',
            mRNA sequence.
ACCESSION   BI560710
VERSION     BI560710.1  GI:15448024
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 779)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11751  row: h  column: 11
            High quality sequence stop: 732.
FEATURES             Location/Qualifiers
     source          1. .779
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5296954"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      234 a    159 c    205 g    181 t
ORIGIN

  Query Match           10.5%;  Score 426.6;  DB 13;  Length 779;
  Best Local Similarity   77.9%;  Pred. No. 5.6e-45;
  Matches  551;  Conservative    0;  Mismatches  151;  Indels    5;  Gaps    3;
```

```
Qy        188 AAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGAGACGGTG 247
              | |||||||||  |||||||||  | ||| | || ||||| || ||||||||||| | |
Db         76 ACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCG 135

Qy        248 CAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACCTCAACAG 307
              ||||  || ||  |  |||| || |||||||||| || || |  || || ||||| | |
Db        136 CAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCG 195

Qy        308 TCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGTCGGAGCC 367
              | ||  |||||||||| ||  ||||||| ||||| ||||| || |    ||  ||||
Db        196 CCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAA---CAGCAGGAGGA 252

Qy        368 ACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACTGGCAGAG 427
              ||| |||||||||||| ||| ||||| || ||||| ||||||||| | ||  | ||||||
Db        253 ACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTAGCAGAG 312

Qy        428 AGAGGAGGGCGTCGCCGAGACTTCCATGACNTTGGTGTGAACACCAGACAGAACCTTGAC 487
              |||||||| |||||  |||| || |||||  |||||||||| || || |||||||| |||
Db        313 AGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGAC 372

Qy        488 CATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCAC 547
              ||||| ||||| |||| ||||| || || ||||| || || | | | |||||  |||||
Db        373 CATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCAT 432

Qy        548 TTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCC 607
              |||||||||| || || || |||||||||| | |||||||| ||||| ||||| || ||
Db        433 TTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCA 492

Qy        608 CTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGA 667
              |||||||| ||| |||| || || || || || || || || || || ||||| || ||||||
Db        493 CTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGA 552

Qy        668 AGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGTC 727
              | ||||||||||||  ||||| || ||||| |||||||||| |||||||||| ||||||
Db        553 AAGTGTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTT 612

Qy        728 ACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCACCCTGACC 787
              || ||||| || ||| |||| |||||||| ||  | |||||||| || |||||  | ||
Db        613 ACTGAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACA 672

Qy        788 AACGAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTTCAGGAGG 847
              || || || ||  | || ||  | || ||  |||||||||||| ||  | |||||||||
Db        673 AATGAACTTCC-ACTACATCAGCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGG 731

Qy        848 CTCTTGAAAATCA-TGAATTTGCAACAAATTGGACGGAATTATTACA 893
              || | ||||||||  ||||||||||||||||||||||  |||||||| |
Db        732 CTTTGGAAAATCATTGAATTTGCAACAAATTGGACGAAATTATTATA 778
```

RESULT 8
BF080923

LOCUS	BF080923 533 bp mRNA linear EST 18-OCT-2000
DEFINITION	233690 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION	BF080923
VERSION	BF080923.1 GI:10874744
KEYWORDS	EST.
SOURCE	pig.
ORGANISM	Sus scrofa Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE	1 (bases 1 to 533)
AUTHORS	Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W. and Keele,J.W.
TITLE	Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
JOURNAL	Unpublished (2000)
COMMENT	Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smith@email.marc.usda.gov Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options. PCR PRimers FORWARD: AGGAAACAGCTATGACCAT BACKWARD: GTTTTCCCAGTCACGACG Plate: 53 row: O column: 21 Seq primer: ATTTAGGTGACACTATAG.

FEATURES	Location/Qualifiers
source	1. .533 /organism="Sus scrofa" /db_xref="taxon:9823" /clone_lib="MARC 2PIG" /tissue_type="pooled" /lab_host="DH10B" /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI; Library made from pooled tissue from testis, ovary, endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 185 a 93 c 119 g 136 t
ORIGIN

```
  Query Match             9.6%;  Score 391.2;  DB 12;  Length 533;
  Best Local Similarity   83.5%;  Pred. No. 2.1e-40;
  Matches  444;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy       1583 CAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGCAGACTGGTCCAAAGAAACAAGA 1642
              || || ||||| || |  |||||||| |||||||||||||| ||||| ||||||||||
Db          2 CAAGGTGGAAAAACATGGGATTACAATCCACAATTTGCAGATTGGTCAAAAGAAACAAGA 61

Qy       1643 GGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGGCTGCTGATCTATACCCGCAGG 1702
              || || ||  | || || || |||||| | |||||||||||| |||||||||| || ||
Db         62 GGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACTCGAAGA 121

Qy       1703 AATTATGAAGCAGCCAACTCACTGATACAGAACCTGTTCAAAGTGACTCCAGCCATGGGC 1762
              |||||||||||||||| ||| ||||||| || || || ||||| || ||||||||||
Db        122 AATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGC 181

Qy       1763 ATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGACAGAACAGAAGCTTATCTGAGA 1822
              || || |||||||| ||||| ||||| || |||||||| ||||| ||||| ||| | |||
Db        182 ATACAAATGAAAAAAGCAATAATGATTGAAGTGGATGATAGAACTGAAGCCTATTTAAGA 241

Qy       1823 GCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGTCTCTTGTCAAGTAAT 1882
              | ||| ||||| ||||| || || || || |||||||||| ||||| ||||||||||
Db        242 GTCTTACAGCAAAAGGTTACATCGGATACCCAGATAGTTGTTTGTCTGTTGTCAAGTAAT 301

Qy       1883 CGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTGTACAGACTGCCCCACCCCAAGT 1942
              || |||||||||||||||| || || || ||||| |||||||| ||||| || ||||||
Db        302 CGAAAGGACAAATATGATGCTATTAAAAAATACTTATGTACAGATTGCCCTACTCCAAGT 361

Qy       1943 CAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATGGCCATTGCCACCAAG 2002
              |||||||||||||||| ||| | ||||| || || || |||||||||||||| || |||
Db        362 CAGTGTGTGGTGGCCCGAACCTTAGGCAAACAACAGACTGTCATGGCCATTGCGACAAAG 421

Qy       2003 ATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGGGTGGACATGGCCCTG 2062
              |||||||||||||||||||||||||||||||| ||||| ||| |||| || ||| | ||
Db        422 ATCGCCCTGCAGATGAACTGCAAGATGGGAGGAGAGCTTTGGAGGGTTGATATGCCTCTA 481

Qy       2063 AAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACACCACAGCTGGGC 2114
              || || || ||||| || ||||| |||| |||||||||||| |
Db        482 AAGCTAGCGATGATAGTTGGCATTGATTGTTATCATGATACCACAGCTGGAC 533
```

RESULT 9
BF664941

LOCUS	BF664941 1495 bp mRNA linear EST 21-DEC-2000
DEFINITION	602119191F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4276247 5', mRNA sequence.
ACCESSION	BF664941
VERSION	BF664941.1 GI:11938836
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1495)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1096  row: f  column: 24
            High quality sequence stop: 511.
FEATURES             Location/Qualifiers
     source          1. .1495
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4276247"
                     /clone_lib="NIH_MGC_56"
                     /tissue_type="primitive neuroectoderm"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA.  5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.65 kb (range 0.9-4.0 kb).  15/15 colonies
                     contained  inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      476 a    280 c    536 g    203 t
ORIGIN

  Query Match             9.5%;  Score 387.2;  DB 12;  Length 1495;
  Best Local Similarity   83.0%;  Pred. No. 3.5e-40;
  Matches  465;  Conservative    0;  Mismatches   93;  Indels    2;  Gaps    2;

Qy       1784 ATGATCGAGGTGGATGACAGAACAGAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACG 1843
              | ||| || |||||||||||||| ||||| ||  | |||| ||| ||||| |||||  ||
Db             8 AGGATTGAAGTGGATGACAGAACTGAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACA 67

Qy       1844 TCAGACACTCAGATAGTTGTCTGTCTCTTGTCAAGTAATCGGAAGGACAAATATGATGCC 1903
               ||||||| |||||||||||||||| ||||||||||||||||||||||||||| |||||
Db            68 GCAGACACCCAGATAGTTGTCTGTCTGTTGTCAAGTAATCGGAAGGACAAATACGATGCT 127

Qy       1904 ATCAAGAAGTACTTGTGTACAGACTGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACC 1963
              || || || ||| |||||||||| ||||| ||||||||||||||||||||||||||| |||
Db           128 ATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACC 187

Qy       1964 CTGGGCAAGCAGCAAACAGTCATGGCCATTGCCACCAAGATCGCCCTGCAGATGAACTGC 2023
              | ||||| |||||||| |||||||||||||| || ||||| ||||| ||||||||||||
Db           188 TTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAAAGATTGCCCTACAGATGAACTGC 247

Qy       2024 AAGATGGGAGGCGAGCTCTGGCGGGTGGACATGGCCCTGAAACTGGCAATGATCGTGGGC 2083
              |||||||||||| ||||||||| |||||||||  ||||||| || |  |||||||| |||
Db           248 AAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGC 307

Qy       2084 ATCGACTGTTACCATGACACCACAGCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGC 2143
              |||||  |||||||||||||  |||||||||||||||| ||||||||||| || ||||||
Db           308 ATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGC 367

Qy       2144 ATCAATGAAGGGATGACCCGCTGGTTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAG 2203
              |||||||||||||||||||||||||||||| |||||| | ||||||||  | || ||||||
Db           368 ATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCATATTTCAGGATAGAGGACAGGAG 427

Qy       2204 CTGGTGGATGGTCTCAAGGTGTGCTTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAAT 2263
              ||||| ||||| ||||| || ||| ||||||| | |||||||||||||| | ||||||||
Db           428 CTGGTAGATGGGCTCAAAGTCTGCCTGCAAGC-GGTCTGAGGGCTTGGAATAGCTGCAAT 486

Qy       2264 GAATACATGCCCAGCCGTGTCATCG-TGTACCGAGACGGTGTGGGGGACGGGCAGCTGAA 2322
              || |||||| ||   ||  |||||| ||||||| || || ||  | || ||     | ||
Db           487 GAGTACATGGCCCAGCGGATCATCGTTGTACCGCGATGGGGTTAGAGAAGGCAGTGTAAA 546

Qy       2323 GACCCTGGTCAATTATGAGG 2342
                |  ||| | |  | | ||
Db           547 CCCGGTGGACTACGAAGGGG 566


RESULT 10
BI463482
LOCUS       BI463482               1092 bp    mRNA    linear   EST 21-AUG-2001
DEFINITION  603203560F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5269311 5',
            mRNA sequence.
ACCESSION   BI463482
VERSION     BI463482.1  GI:15254125
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1092)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11679  row: h  column: 16
            High quality sequence stop: 668.
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5269311"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      356 a    201 c    270 g    264 t      1 others
ORIGIN

  Query Match             9.4%;  Score 383.6;  DB 13;  Length 1092;
  Best Local Similarity   74.3%;  Pred. No. 1.2e-39;
  Matches  551;  Conservative    0;  Mismatches  181;  Indels   10;  Gaps    5;

Qy        197 GGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGAGACGGTGCAGCATGTT 256
              ||   |||||||||  | ||| | || ||||| || ||||||||||| | ||||  || |
Db         83 GGAGAAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGACTGGT 142

Qy        257 GGGG-CTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACCTCAACAGTCCCCCAC 315
              |||  |  |||| || |||||||||| || || |  || || ||||| | | | ||  |
Db        143 GGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGCCACCAGC 202

Qy        316 AGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGTCGGAGCCACATCCAA 375
              ||||||||| ||  ||||||| ||||| ||||| || |    ||  ||||  ||| ||||
Db        203 AGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAA---CAGCAGGAGGAACAGCCAA 259
```

```
Qy        376 GTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACTGGCAGAGAGAGGAGG 435
Db        260 GTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTAGCAGAGAGAGGAGG 319

Qy        436 GCGTCGCCGAGACTTCCATGACNTTGGTGTGAACACCAGACAGAACCTTGACCATGTCAA 495
Db        320 TCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGTTAA 379

Qy        496 AGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCACTTCCGGCT 555
Db        380 AGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCT 439

Qy        556 GACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCCCTGATGGA 615
Db        440 GACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGGA 499

Qy        616 GGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGAAGGTGTCA 675
Db        500 AGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCA 559

Qy        676 TGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGTCACAGAAGT 735
Db        560 TGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAAGGTTACTGAAG 619

Qy        736 ATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCACCCTGACCAACGAGCT 795
Db        620 TTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCAACTTTAACAAATGATC 679

Qy        796 GCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTTCAGGAGGCTCTTGAA 855
Db        680 TTCAACTACAT--ACCAACTTGTTGAAAGTCTATAATATTATTTCAGGGAGGCT-TTGAA 736

Qy        856 AATCATGAATTTGCAACAAATTGGACGGAATTATTACAATC---CAAGTGACCCGATTGA 912
Db        737 AATCATGAATTTGAAACAAATTGGACGAAATTATTATAAGCACACAATGAACCGAGTTGA 796

Qy        913 TATTCCAAACCACAGGTTGGTG 934
Db        797 TATTACAAGTAACAGGGTGGGG 818
```

RESULT 11
BB709598

LOCUS	BB709598 422 bp mRNA linear EST 11-OCT-2001
DEFINITION	BB709598 RIKEN full-length enriched, 2 cells egg Mus musculus cDNA clone B020003D11 3', mRNA sequence.
ACCESSION	BB709598
VERSION	BB709598.1 GI:16062767
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 422)
AUTHORS	Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE	RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
JOURNAL	Unpublished (2001)
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000) wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000) Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001) Please visit our web site (http://genome.gsc.riken.go.jp) for further details. e mouse tissues.

FEATURES Location/Qualifiers

source	1. .422 /organism="Mus musculus" /db_xref="taxon:10090" /clone="B020003D11" /clone_lib="RIKEN full-length enriched, 2 cells egg" /tissue_type="egg" /dev_stage="2 cells" /lab_host="DH10B" /note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 80 a 93 c 108 g 141 t
ORIGIN

Query Match 9.1%; Score 370.8; DB 10; Length 422;
Best Local Similarity 96.5%; Pred. No. 9e-38;
Matches 413; Conservative 0; Mismatches 7; Indels 8; Gaps 3;

```
Qy       2857 TTCTAGCGGGATGTTGGGGGTAGGGGTTGGGGGCCGCTGTTTTGTTTTGTTTTGTTTTTA 2916
Db          2 TTCTAACGGGATGTT-GGGGTAGGGGTTGGGGCCCCCTCTTTTCTTTTGTTTTGTTTTTA 60

Qy       2917 GTTGTTGGTGTGTAGCGGCTATAGAGTTAACTAAGGCTTAT-GTTTGCTTTGTTTCTCAT 2975
Db         61 GTTCTTGGTGTGTAGCGGCTATAGAGTTAACTAAGGCTTATAGTTTGCTTTGTTTCTCAT 120

Qy       2976 GGTGTGTTGTGACAGTTTGCTTGGCAGAGTGGTGGCAAGGGCCGTGGTACCACAGATAGA 3035
Db         121 GGTGTGTTGTGACAGTTTGCTTGGCAGAGTGGTGGCAAGGGCCGTGGTACCACAGATAGA 180

Qy       3036 CTTTCCTAAAGCAGCCGTGTTGGCAGGTTCTTGGAGGTGCTCCTGGGAGCACCTCCTATT 3095
Db         181 CTTTCCTAAAGCAGCCGTGTTGGCAGGTTCTTGGAGGTGCTCCTGGGAGCACCTCCTATT 240

Qy       3096 GTACGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3155
Db         241 GTACCCTG------CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 294

Qy       3156 CTGCATGTTTGAATTTAAAAGAGGAAAGGACACAAGGCAAGAATGCCCAGAGCATCCTAG 3215
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 CTGCATGTTTGAATTTAAAAGAGGAAAGGACACAAGGCAAGAATGCCCAGAGCATCCTAG 354

Qy       3216 ATGGGTGCTATATGGGACTCTGACACCTCAGGTTCTTAAAGATAAAAAGGCCAGAGGTAC 3275
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        355 ATGGGTGCTATATGGGACTCTGACACCTCAGGTTCTTAAAGATAAAAAGGCCAGAGGTAC 414

Qy       3276 CTTTCTGT 3283
                ||||||||
Db        415 CTTTCTGT 422


RESULT 12
BI459464
LOCUS       BI459464                882 bp    mRNA    linear   EST 21-AUG-2001
DEFINITION  603200456F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5266428 5',
            mRNA sequence.
ACCESSION   BI459464
VERSION     BI459464.1  GI:15250120
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 882)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11671  row: p  column: 13
            High quality sequence stop: 717.
FEATURES             Location/Qualifiers
     source          1. .882
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5266428"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      252 a    196 c    231 g    203 t
ORIGIN

  Query Match             9.1%;  Score 370;  DB 13;  Length 882;
  Best Local Similarity   72.4%;  Pred. No. 7.1e-38;
  Matches  575;  Conservative    0;  Mismatches  207;  Indels   12;  Gaps    7;

Qy        188 AAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGAGACGGTG 247
              | |||||||||  ||||||||||  | ||| | || ||||| || ||||||||||| | |
Db         93 ACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCG 152

Qy        248 CAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACCTCAACAG 307
              ||||  || ||  |  |||| || ||||||||||| || || |  || || ||||| | |
Db        153 CAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCG 212

Qy        308 TCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGTCGGAGCC 367
                | ||  |||||||||| ||  ||||||| ||||| ||||| || |    ||  ||||

Db        213 CCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAA---CAGCAGGAGGA 269

Qy        368 ACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACTGGCAGAG 427
              ||| |||||||||||| ||| ||||| || ||||| ||||||||| | ||  | ||||||
Db        270 ACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTAGCAGAG 329

Qy        428 AGAGGAGGGCGTCGCCGAGACTTCCATGACNTTGGTGTGAACACCAGACAGAACCTTGAC 487
              |||||||| |||||  |||| || |||||  |||||||||| || || |||||||| |||
Db        330 AGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGAC 389

Qy        488 CATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCAC 547
              ||||| ||||| ||||| ||||| || || ||||| || || | | | ||||| |||||
Db        390 CATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCAT 449

Qy        548 TTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCC 607
              |||||||||||| || || || |||||||| | |||||||| ||||| ||||| || ||
Db        450 TTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAA-CCA 508

Qy        608 CTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGA 667
              |||||||| ||| |||| || || || || || || || || || || ||||| || ||||||
Db        509 CTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGA 568

Qy        668 AGGTGTCATGCTTTCGATGG-GACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGT 726
              | |||||||||||||| |||||  || ||||| |||||||||||| |||||||||| |||||
Db        569 AAGTGTCATGCTTTTGATGGAAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGT 628

Qy        727 CACAGAAGTATTCAGTCAGACTCGGAATGGGGAA-CACGTGAGGATCACC-ATCACCCTG 784
              || ||||| || ||| |||  |||||||| |||  | |||||||| ||| |||||  |
Db        629 TACTGAAGTTTTTAGTAAGAACCGGAATGGAGAAGGATGTGAGGATTACCGATCACTTTA 688

Qy        785 ACCAACGAGCTGCCGCCCACCTCGCCCA----CCTGCCTGCAGTTCTATAACATCNTCTT 840
              |  | || || || |  |  |  |||       |        |||||||||| ||  | |
Db        689 CCACATGAACTTCCACGTAAATTCACCAAACTTGTTTGCCAGGTTCTATAATATTATTTC 748

Qy        841 CAGGAGGCTCTTGAAAATCATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAG 900
              ||| ||| | || | ||  || ||||| || || ||| || || | ||| |  || |
Db        749 ACGGAAGCTTTGGACACTCCCGACTTTGCCACCAAATGGCCGAAAATTTTATATACCGAA 808

Qy        901 TGACCCGATTGATATTCCAAACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCAT 960
              ||  || |||||  |||| |  ||| |||||||||  |||||  |  |      |  |||
Db        809 TGGACCAATTGAGTTTCC-AGTCACCGGTTGGTGAATTGGCCGGATTCACTAAATGCCAC 867

Qy        961 CCTTCAGTATGAGA 974
              | |  ||  || |
Db        868 GCCTAGGTTGGACA 881


RESULT 13
AL705072
LOCUS       AL705072                622 bp    mRNA    linear   EST 22-MAR-2002
DEFINITION  DKFZp686E0634_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
            DKFZp686E0634 5', mRNA sequence.
ACCESSION   AL705072
VERSION     AL705072.1  GI:19688427
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 622)
  AUTHORS   Ottenwaelder,B., Obermaier,B., Mewes,W., Mewes,H.W., Weil,B. and
            Wiemann,S.
  TITLE     EST (Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and Wiemann
            ,S.)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Ottenwaelder B
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by MediGenomix (Martinsried/Germany) within the cDNA
            sequencing consortium of the German Genome Project. No s1 sequence
```

```
            available.
            This clone (DKFZp686E0634) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .622
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686E0634"
                     /clone_lib="686 (synonym: hlcc3)"
                     /tissue_type="human skeletal muscle"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
BASE COUNT      187 a    134 c    168 g    133 t
ORIGIN

  Query Match             9.0%;  Score 364.6;  DB 9;  Length 622;
  Best Local Similarity   77.9%;  Pred. No. 4.3e-37;
  Matches  452;  Conservative    0;  Mismatches  125;  Indels    3;  Gaps    1;

Qy        188 AAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGAGACGGTG 247
              | |||||||||  ||||||||||  | ||| | || ||||| || ||||||||||| | |
Db         46 ACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCG 105

Qy        248 CAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACCTCAACAG 307
              ||||  || ||  |  |||| || |||||||||||| || || |  || || |||||  | |
Db        106 CAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCG 165

Qy        308 TCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGTCGGAGCC 367
                | ||  ||||||||||| ||  ||||||| ||||| ||||| || |     ||  ||||
Db        166 CCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAA---CAGCAGGAGGA 222

Qy        368 ACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACTGGCAGAG 427
              ||| ||||||||||||| ||| ||||| || ||||| ||||||||| | ||  | ||||||
Db        223 ACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTAGCAGAG 282

Qy        428 AGAGGAGGGCGTCGCCGAGACTTCCATGACNTTGGTGTGAACACCAGACAGAACCTTGAC 487
              |||||||| |||||  |||| || |||||  |||||||||| || || |||||||| |||
Db        283 AGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGAC 342

Qy        488 CATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCAC 547
              ||||| ||||| ||||| ||||| || || ||||| || || | | | ||||| |||||
Db        343 CATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCAT 402

Qy        548 TTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCC 607
              |||||||||||  || || || |||||||| | |||||||| ||||| ||||| || ||
Db        403 TTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCA 462

Qy        608 CTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGA 667
              |||||||| ||| |||| || || || || || || || || || ||||| || ||||||
Db        463 CTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGA 522

Qy        668 AGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGTC 727
              | ||||||||||| ||||| || ||||| |||||||||| |||||||||| |||||
Db        523 AAGTGTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTT 582

Qy        728 ACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGA 767
              || ||||| || ||| |||| |||||||| ||  | ||||
Db        583 ACTGAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGA 622


RESULT 14
BM751055
LOCUS       BM751055                 527 bp    mRNA    linear   EST 04-MAR-2002
DEFINITION  K-EST0026956 S9SNU601 Homo sapiens cDNA clone S9SNU601-21-E08 5',
            mRNA sequence.
ACCESSION   BM751055
VERSION     BM751055.1  GI:19080673
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 527)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 21  row: E  column: 08
            High quality sequence stop: 527.
FEATURES             Location/Qualifiers
     source          1. .527
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="S9SNU601-21-E08"
                     /clone_lib="S9SNU601"
                     /sex="M"
                     /tissue_type="Ascites"
                     /cell_type="Epithelial"
                     /cell_line="SNU-601"
                     /lab_host="Top10F'"
                     /note="Organ: Stomach; Vector: pME18-FL3; Site_1: XhoI;
                     Site_2: XhoI; The poly (A)+ RNA was dephosphorylated with
                     bacterial alkaline phosphatase (BAP) and then decapped
                     with tabacco acid pyrophosphatase (TAP). The decapped
                     intact mRNA was ligated with DNA-RNA linker including SfiI
                     site by treatment of T4 RNA ligase and the first strand
                     cDNA was synthesized with Superscript II using SfiI
                     oligo-dT primer. After first strand synthesis, RNA was
                     degraded by NaOH treatment and cDNA was amplified by PCR
                     reaction. The PCR products were digested with SfiI and
                     cloned into DraIII- digested pME18S-FL3 vector. The
                     obtained cDNA vectors were used for transformation of
                     competent cells E. coli Top10F' by electroporation method.
                     The cDNA libraries constructed by this method are
                     full-length enriched cDNA library."
BASE COUNT      188 a     95 c    113 g    131 t
ORIGIN

  Query Match             8.9%;  Score 363.6;  DB 14;  Length 527;
  Best Local Similarity   81.9%;  Pred. No. 6.4e-37;
  Matches  431;  Conservative    0;  Mismatches   94;  Indels    1;  Gaps    1;

Qy       1381 GCGCAATGATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCA 1440
              ||| |||||||| || |||||||| ||| | || || || ||  | || || || |||||
Db          1 GCGTAATGATTTTAACGTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCA 60

Qy       1441 GCGGCAGCGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAG 1500
              ||||||| || ||||| || ||||| || |||||||| ||  | || ||||| || ||
Db         61 AAGGCAGCGTGAAGTGGGACGACTCATTGATTACATCCATAAAAACGATAATGTTCAAAG 120

Qy       1501 AGAGCTTCGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAAT 1560
              ||||||||||||||||  ||||||| || || ||||| |||||||||| ||||||||
Db        121 GGAGCTTCGAGACTGGGGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAAT 180

Qy       1561 CTTACAATCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGC 1620
              || ||| | || ||||| ||||| || ||||| || |||||||||| ||||||||||
Db        181 TTTGCAAACAGAAAAGATTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGC 240

Qy       1621 AGACTGGTCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTG 1680
              ||| ||||||||||||||||||| || ||  | || || || |||||| | ||||||||
Db        241 AGATTGGTCCAAAGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTG 300

Qy       1681 GCTGCTGATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTT 1740
              |||| |||||||||| || || |||||||||||||||| ||| ||||||| || || ||
Db        301 GCTGTTGATCTATACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATT 360
```

```
Qy       1741 CAAAGTGACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGA 1800
             ||||| || ||||||||||||||| || |||| ||| ||||| ||||| || ||||||||
Db        361 TAAAGTTACACCAGCCATGGGCATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGA 420

Qy       1801 CAGAACAGAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGT 1860
             |||||| ||||| ||  | |||| ||| ||||| ||||| ||  ||||||| ||||||||
Db        421 CAGAACTGAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGT 480

Qy       1861 TGTCTGTCTCTTGTCAAGTAATC-GGAAGGACAAATATGATGCCAT 1905
             |||||||| |||||||||||||| |||||||||||| ||||| ||
Db        481 TGTCTGTCTGTTGTCAAGTAATCGGGAAGGACAAATACGATGCTAT 526


RESULT 15
BI463897
LOCUS       BI463897                 666 bp    mRNA    linear   EST 21-AUG-2001
DEFINITION  603202635F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5268465 5',
            mRNA sequence.
ACCESSION   BI463897
VERSION     BI463897.1  GI:15254553
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 666)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11677  row: e  column: 10
            High quality sequence stop: 666.
FEATURES             Location/Qualifiers
     source          1. .666
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5268465"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      194 a    144 c    187 g    141 t
ORIGIN

  Query Match             8.9%;  Score 363.4;  DB 13;  Length 666;
  Best Local Similarity   77.9%;  Pred. No. 5.8e-37;
  Matches  463;  Conservative    0;  Mismatches  127;  Indels    4;  Gaps    2;

Qy        188 AAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGAGACGGTG 247
              | |||||||||  ||||||||||  | ||| | || ||||| || ||||||||||| | |
Db         76 ACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCG 135

Qy        248 CAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACCTCAACAG 307
              ||||  || ||   |  |||| || |||||||||||| || || |  || || ||||| | |
Db        136 CAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCG 195

Qy        308 TCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGTCGGAGCC 367
              | ||  |||||||||| ||  ||||||| ||||| ||||| || |    ||  ||||
Db        196 CCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAA---CAGCAGGAGGA 252

Qy        368 ACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACTGGCAGAG 427
              ||| |||||||||||| ||| ||||| || ||||| |||||||| | ||  | ||||||
Db        253 ACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTAGCAGAG 312

Qy        428 AGAGGAGGGCGTCGCCGAGACTTCCATGACNTTGGTGTGAACACCAGACAGAACCTTGAC 487
              |||||||| |||||  |||| || |||||  |||||||||| || || |||||||| |||
Db        313 AGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGAC 372

Qy        488 CATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCAC 547
              ||||| ||||| ||||| ||||| || || ||||| || || | | | ||||| |||||
Db        373 CATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCAT 432

Qy        548 TTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCC 607
              |||||||||| || || || ||||||||| | |||||||| ||||| ||||| || ||
Db        433 TTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCA 492

Qy        608 CTGATGG-AGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGG 666
              ||||||| | ||| |||| || || || || || || || || ||||| || |||||
Db        493 CTGATGGAACGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGG 552

Qy        667 AAGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGT 726
              || |||||||||||| ||||| || ||||| |||||||||| |||||||||| |||||
Db        553 AAAGTGTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGT 612

Qy        727 CACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCAC 780
              || ||||| || ||| |||| |||||||| ||  | |||||||| || |||||
Db        613 TACTGAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCAC 666


Search completed: July 21, 2003, 22:53:03
Job time : 5341 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2003, 19:01:27 ; Search time 30.2338 Seconds
(without alignments)
3799.123 Million cell updates/sec

Title: US-09-873-737A-4
Perfect score: 4528
Sequence: 1 MTGRARARARGRARGQETVQ.....VGQSIHREPNLSLSNRLYYL 862

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4520	99.8	862	21	AAY90234	Mouse piwi protein
2	4286.5	94.7	861	21	AAY90235	Human piwi protein
3	1459.5	32.2	866	22	ABB62102	Drosophila melanog
4	1398.5	30.9	843	22	ABB62084	Drosophila melanog
5	1385.5	30.6	843	21	AAY90233	Drosophila piwi pr
6	1360.5	30.0	580	22	AAU07866	Polypeptide sequen
7	1190	26.3	530	22	AAB94209	Human protein sequ
8	932	20.6	371	22	AAB92722	Human protein sequ
9	620	13.7	130	21	AAB58977	Breast and ovarian
10	579	12.8	341	22	ABG05957	Novel human diagno
11	506.5	11.2	983	21	AAG51562	Arabidopsis thalia
12	506.5	11.2	988	21	AAG51561	Arabidopsis thalia
13	500	11.0	906	21	AAG51563	Arabidopsis thalia
14	499.5	11.0	1048	21	AAG42118	Arabidopsis thalia
15	493.5	10.9	1064	21	AAG52593	Arabidopsis thalia
16	485	10.7	868	21	AAG42119	Arabidopsis thalia
17	484	10.7	860	22	AAM80292	Human protein SEQ
18	484	10.7	860	22	AAB94291	Human protein sequ
19	479	10.6	884	21	AAG52594	Arabidopsis thalia
20	477	10.5	950	22	ABB67046	Drosophila melanog
21	477	10.5	984	22	ABB62573	Drosophila melanog
22	477	10.5	984	22	ABB67045	Drosophila melanog
23	460.5	10.2	836	21	AAG52595	Arabidopsis thalia
24	458.5	10.1	782	22	AAB93139	Human protein sequ
25	458.5	10.1	820	21	AAG42120	Arabidopsis thalia
26	422.5	9.3	1145	22	AAE09770	Drosophila melanog
27	413	9.1	962	22	ABB63278	Drosophila melanog
28	356	7.9	938	22	AAB82553	Neurospora crassa
29	338.5	7.5	199	22	ABG05956	Novel human diagno
30	313	6.9	108	22	AAO04945	Human polypeptide
31	306.5	6.8	326	21	AAG52610	Arabidopsis thalia
32	304	6.7	1020	22	AAU01856	C. elegans RNA int
33	297	6.6	140	22	ABG12914	Novel human diagno
34	294.5	6.5	406	21	AAG54509	Zea mays protein f
35	279.5	6.2	227	21	AAG52611	Arabidopsis thalia
36	276	6.1	252	22	AAB97112	Translation initia
37	248.5	5.5	352	21	AAG54511	Zea mays protein f
38	248.5	5.5	353	21	AAG54510	Zea mays protein f
39	210.5	4.6	71	23	ABG40429	Human peptide enco
40	208	4.6	232	22	AAM85465	Human immune/haema
41	194.5	4.3	747	21	AAG25621	Arabidopsis thalia
42	194.5	4.3	749	21	AAG25620	Arabidopsis thalia
43	180	4.0	567	21	AAG25622	Arabidopsis thalia
44	151.5	3.3	1924	22	AAE09768	Human dicer protei
45	144	3.2	55	22	ABB37543	Peptide #5049 enco

ALIGNMENTS

RESULT 1
AAY90234
ID AAY90234 standard; Protein; 862 AA.
XX
AC AAY90234;
XX
DT 29-AUG-2000 (first entry)
XX
DE Mouse piwi protein, designated miwi.
XX
KW Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW anaemia; immunodeficiency; male infertility; mouse; ds.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT Misc-difference 90
FT /label= Leu, Ile
FT /note= "encoded by NTT"
FT Misc-difference 216
FT /note= "unspecified amino acid; encoded by NTC"
FT Misc-difference 383
FT /label= Leu, Ile
FT /note= "encoded by NTC"
FT Misc-difference 816
FT /label= Leu, Ile
FT /note= "encoded by NTC"
XX
PN WO200032039-A1.
XX
PD 08-JUN-2000.
XX
PF 03-DEC-1999; 99WO-US28764.

```
XX
PR   04-DEC-1998;   98US-0110901.
XX
PA   (UYDU-) UNIV DUKE.
XX
PI   Lin H;
XX
DR   WPI; 2000-412085/35.
DR   N-PSDB; AAA07587.
XX
PT   Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT   therapy of diseases such as cancer and in various research and
PT   diagnostic applications -
XX
PS   Claim 4; Page 180-185; 201pp; English.
XX
CC   This sequence represents the mouse piwi family protein, designated
CC   miwi. The piwi family nucleic acids and polypeptides are used in gene
CC   therapy of diseases such as cancer and also in various research and
CC   diagnostic applications. The sequences can also be used to treat
CC   tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ   Sequence   862 AA;

  Query Match             99.8%;  Score 4520;  DB 21;  Length 862;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  862;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEGDLVGRGRQRGMVVGAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEGDLVGRGRQRGMVVGAT 60

Qy         61 SKSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHF 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHF 120

Qy        121 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVT 180

Qy        181 EVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 EVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPI 240

Qy        241 DIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 DIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQV 300

Qy        301 SKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQ 360

Qy        361 PVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 PVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 420

Qy        421 EVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWS 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWS 480

Qy        481 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTE 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTE 540

Qy        541 AYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 600

Qy        601 IATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWF 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWF 660

Qy        661 SRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQ 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 SRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQ 720

Qy        721 FLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 FLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 780

Qy        781 VSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 VSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLA 840

Qy        841 FLVGQSIHREPNLSLSNRLYYL 862
              ||||||||||||||||||||||
Db        841 FLVGQSIHREPNLSLSNRLYYL 862


RESULT 2
AAY90235
ID   AAY90235 standard; Protein; 861 AA.
XX
AC   AAY90235;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Human piwi protein, designated hiwi.
XX
KW   Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW   anaemia; immunodeficiency; male infertility; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 76
FT                   /label= Leu, Ile
FT                   /note= "encoded by NTA"
FT   Misc-difference 303
FT                   /label= Leu, Ile
FT                   /note= "encoded by NTA"
FT   Misc-difference 735
FT                   /label= Leu, Ile
FT                   /note= "encoded by NTA"
XX
PN   WO200032039-A1.
XX
PD   08-JUN-2000.
XX
PF   03-DEC-1999;   99WO-US28764.
XX
PR   04-DEC-1998;   98US-0110901.
XX
PA   (UYDU-) UNIV DUKE.
XX
PI   Lin H;
XX
DR   WPI; 2000-412085/35.
DR   N-PSDB; AAA07588.
XX
PT   Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT   therapy of diseases such as cancer and in various research and
PT   diagnostic applications -
XX
PS   Claim 4; Page 189-194; 201pp; English.
XX
CC   This sequence represents the human piwi family protein, designated
CC   hiwi. The piwi family nucleic acids and polypeptides are used in gene
CC   therapy of diseases such as cancer and also in various research and
CC   diagnostic applications. The sequences can also be used to treat
CC   tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ   Sequence   861 AA;

  Query Match             94.7%;  Score 4286.5;  DB 21;  Length 861;
  Best Local Similarity   93.7%;  Pred. No. 0;
```

```
  Matches  808;  Conservative   22;  Mismatches   31;  Indels    1;  Gaps    1;

Qy          1 MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEGDLVGRGRQRGMVVGAT 60
              |||||||||||||||||| | ||: |||||||| ||||  | ||:| |||||||   | |
Db          1 MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRG-TAGGT 59

Qy         61 SKSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHF 120
              :||| |||||||||||| |||||||||||| ||||||||||||||||||||||:||||||
Db         60 AKSQGLQISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHF 119

Qy        121 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVT 180
              |||||||||||||||||||||||||||||||||||||:|||||||||||||||| |||
Db        120 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVT 179

Qy        181 EVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPI 240
              ||||:|||| |||||||||||||||||||| ||||||||||||||||||:|||
Db        180 EVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPI 239

Qy        241 DIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQV 300
              |||:|||||||||||||||:|||||||||||||||||||||| | ||||||||||
Db        240 DIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQV 299

Qy        301 SKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQ 360
              ||| |||:|||||| |||||||||||||||||||||| ||||||||||||||||||||
Db        300 SKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQ 359

Qy        361 PVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 420
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        360 PVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 419

Qy        421 EVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWS 480
              ||||||||||:|||||||||||||||||||||||||||||:|||||||||||||||||||
Db        420 EVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWS 479

Qy        481 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTE 540
              |||||||||||||||||||||||||||||||||||||||||||:||:||||||||||||
Db        480 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDRTE 539

Qy        541 AYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 600
              |||| ||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 AYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 599

Qy        601 IATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWF 660
              |||||||||||||||||||: ||| |||||||||| ||||||||||||||||||||||||
Db        600 IATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWF 659

Qy        661 SRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQ 720
              |||:|||||||||||||||||||||: |||||||:|||||||||||||||||||||||||
Db        660 SRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQ 719

Qy        721 FLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 780
              |||||||:|||||||| |||||||||| ||||||||||||||||||||||||||||||||
Db        720 FLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 779

Qy        781 VSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLA 840
              |||||||||||||||||:|||||||||||||| ||:|||||||||||||||||||||||||
Db        780 VSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLA 839

Qy        841 FLVGQSIHREPNLSLSNRLYYL 862
              ||||||||||||||||||||||
Db        840 FLVGQSIHREPNLSLSNRLYYL 861


RESULT 3
ABB62102
ID   ABB62102 standard; Protein; 866 AA.
XX
AC   ABB62102;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 13098.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL06205.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 13098; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   866 AA;

  Query Match             32.2%;  Score 1459.5;  DB 22;  Length 866;
  Best Local Similarity   36.4%;  Pred. No. 4.2e-136;
  Matches  320;  Conservative  174;  Mismatches  339;  Indels   47;  Gaps   18;

Qy          7 ARARGRARGQETVQHVGAAASQQPGYIPPRPQQS-PTEGDLVGRGRQRGMVVGATSKSQE 65
              || ||| |    |:          |: |   |:| |:  :    | |     |      :
Db         10 ARGRGRGRKPNNVE-------ANRGFAPSLGQKSDPSHSE----GNQAS---GGNGGGGD 55

Qy         66 LQISAGFQELSLAERGGRRRDFHDXG-VNTRQNLDHV-------KESKTGSSGIIVKLST 117
               |:    :: ||:     : :    | |  |: :  :              || |  |  : :
Db         56 AQVGPSIEKSSLSAVQMHKSEGDPRGSVRGRRLITDLVYSRPPGMTSKKGVVGTHITVQA 115

Qy        118 NHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLP---KR 174
              |:|::  || | :||| :|: | :|| ||| : |::|: ::|  : |||| :|      |
Db        116 NYFKVLKRPNWTIYQYRVDFTPDVEATRLRRSFLYEHKGILGG-YIFDGTNMFCINQFKA 174

Qy        175 LQHK--VTEVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRN 232
              :|    | |: :::| ||:: | |     : |     |  |:  || :: ::|: : |
Db        175 VQDSPYVLELVTKSRAGENIEIKIKAVGSVQSTDAEQFQVLNLILRRAMEGLDLKLVSRY 234

Qy        233 YYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTE 292
              ||:|   |:: | |: :|||: ||| |:||:|:||::: |||:|:||: : : :  : ::
Db        235 YYDPQAKINLENFRMQLWPGYQTSIRQHENDILLCSEICHKVMRTETLYNILSDAIRDSD 294

Qy        293 EHKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYN 352
              :  :|    : ::|:::|| |||||||:||:|:   |   ||  || |:|:::||:|:||
Db        295 D--YQSTFKRAVMGMVILTDYNNKTYRIDDVDFQSTPLCKFKTNDG-EISYVDYYKKRYN 351

Qy        353 QEITDLKQPVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTR 412
              | |||||:::|:|  :  ||      | ||||   ||:|| || ||  :: :: |||
Db        352 IIIRDLKQPLVMSRPTDKNIRGGN-DQAIMIIPELARATGMTDAMRADFRTLRAMSEHTR 410
```

```
Qy        413 LTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDY 472
Db        411 LNPDRRIERLRMFNKRLKSCKQSVETLKSWNIELDSALVEIPARVLPPEKILFGNQKIFV 470

Qy        473 NPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIM 532
Db        471 CDARADWTNEFRTCSMFKNVHINRWYVITPSRNLRETQEFVQMCIRTASSMKMNICNPIY 530

Qy        533 IEV-DDRTEAYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVART 591
Db        531 EEIPDDRNGTYSQAIDNAAANDPQIVMVVMRSPNEEKYSCIKKRTCVDRPVPSQVVTLKV 590

Qy        592 LGKQQT----VMAIATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAG 647
Db        591 IAPRQQKPTGLMSIATKVVIQMNAKLMGAPWQVVIPLHGLMTVGFDVCHSPKNKNKSYGA 650

Qy        648 FVASINEGMT-RWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVG 706
Db        651 FVATMDQKESFRYFSTVNEHIKGQELSEQMSVNMACALRSYQEQHRSLPERILFFRDGVG 710

Qy        707 DGQLKTLVNYEVPQFLDCL----KSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLP 762
Db        711 DGQLYQVVNSEVNTLKDRLDEIYKSAGKQEGCRMTFIIVSKRINSRYF--TGHR--NPVP 766

Qy        763 GTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYN 822
Db        767 GTVVDDVITLPERYDFFLVSQAVRIGTVSPTSYNVISDNMGLNADKLQMLSYKMTHMYYN 826

Qy        823 WPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 862
Db        827 YSGTIRVPAVCHYAHKLAFLVAESINRAPSAGLQNQLYFL 866


RESULT 4
ABB62084
ID   ABB62084 standard; Protein; 843 AA.
XX
AC   ABB62084;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 13044.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL06187.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 13044; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   843 AA;

  Query Match             30.9%;  Score 1398.5;  DB 22;  Length 843;
  Best Local Similarity   35.9%;  Pred. No. 5.1e-130;
  Matches  307;  Conservative  172;  Mismatches  318;  Indels   59;  Gaps   18;

Qy         48 GRGRQRGMVVGATSKSQ--------------------ELQISAGFQELSLAE----RGG 82
Db          6 GRGRRRPLNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG 65

Qy         83 --RRRDFHD--XGVNTRQNLDHVKESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYN 138
Db         66 PTERKPWGDQYDYLNTR---PAELVSKKGTDGVPVMLQTNFFRLKTKPEWRIVHYHVEFE 122

Qy        139 PLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVTEVFSQTRNGEHVRITITL 198
Db        123 PSIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIKF 182

Qy        199 TNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSIL 258
Db        183 VGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSIR 242

Qy        259 QYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTY 318
Db        243 QHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRTY 300

Qy        319 RVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLP 378
Db        301 RINDVDFGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNAS 358

Qy        379 GPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQ-- 436
Db        359 ELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPES 415

Qy        437 -RELRDWGLSFDSNLLSFSGRILQSEKI--HQGGKTFDYNPQFADWSKETRGAPLISVKP 493
Db        416 VKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTPS 472

Qy        494 --LDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKK-AIMIEVDDRTEAYLRALQQKV 550
Db        473 DGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCV 532

Qy        551 TSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMN 610
Db        533 RSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQLN 591

Qy        611 CKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASIN-EGMTRWFSRCVFQDRG 669
Db        592 CKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAF 651

Qy        670 QELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVG 729
Db        652 DVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEY 711

Qy        730 RGYN---PRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVR 786
Db        712 ARVQLSPPQLAYIVVTRSMNTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVSQQVR 767

Qy        787 SGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVGQS 846
Db        768 QGTVSPTSYNVLYSSMGLSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTN 827

Qy        847 IHREPNLSLSNRLYYL 862
Db
```

```
Db        828 LHSIPQNALEKKFYYL 843


RESULT 5
AAY90233
ID   AAY90233 standard; Protein; 843 AA.
XX
AC   AAY90233;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Drosophila piwi protein.
XX
KW   Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW   anaemia; immunodeficiency; male infertility; Drosophila.
XX
OS   Drosophila sp.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 13
FT                   /label= Leu, Ile
FT                   /note= "encoded by NTT"
FT   Misc-difference 106
FT                   /label= Leu, Ile
FT                   /note= "encoded by NTA"
FT   Misc-difference 785
FT                   /label= Leu, Ile
FT                   /note= "encoded by NTC"
XX
PN   WO200032039-A1.
XX
PD   08-JUN-2000.
XX
PF   03-DEC-1999;   99WO-US28764.
XX
PR   04-DEC-1998;   98US-0110901.
XX
PA   (UYDU-) UNIV DUKE.
XX
PI   Lin H;
XX
DR   WPI; 2000-412085/35.
DR   N-PSDB; AAA07586.
XX
PT   Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT   therapy of diseases such as cancer and in various research and
PT   diagnostic applications -
XX
PS   Claim 4; Page 171-175; 201pp; English.
XX
CC   This sequence represents the Drosophila piwi family protein, designated
CC   piwi. The piwi family nucleic acids and polypeptides are used in gene
CC   therapy of diseases such as cancer and also in various research and
CC   diagnostic applications. The sequences can also be used to treat
CC   tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ   Sequence   843 AA;

  Query Match           30.6%;  Score 1385.5;  DB 21;  Length 843;
  Best Local Similarity  35.7%;  Pred. No. 1e-128;
  Matches  306;  Conservative  173;  Mismatches  318;  Indels   59;  Gaps   19;

Qy         48 GRGRQRGMVVGATSKSQ--------------------ELQISAGFQELSLAE----RGG 82
              ||||:|      :| |:                    : :| |  :  :| |     ||
Db          6 GRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG 65

Qy         83 --RRRDFHD--XGVNTRQNLDHVKESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYN 138
                 |: : |    :|||  :: |  || |: |: | | || ||  ::|:| :  ||:::
Db         66 PPERKPWGDQYDYLNTRP-VELV--SKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFV 122

Qy        139 PLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVTEVFSQTRNGEHVRITITL 198
              | :|  |:|  :|  | :|:|  : |||  ||  :: : ::| :  :::     :|:|
Db        123 PTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIKF 182

Qy        199 TNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSIL 258
                 :    |  |:|  |:  || :| :||: :||| ::|   |:|   :: :|||: |||
Db        183 VGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSIR 242

Qy        259 QYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTY 318
              |:| :|:| |:::|||:|:||: | |          |  |::|   :: ||||| |||:||
Db        243 QHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRTY 300

Qy        319 RVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLP 378
              |::|:|: | |||||    | ::||:|||  :||  | |  ||:|:|:  | :
Db        301 RINDVDFGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNAS 358

Qy        379 GPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQ-- 436
                : ||||| :|||  :||::| :|: :: :||: |:||     ||  : |:  |
Db        359 ELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPES 415

Qy        437 -RELRDWGLSFDSNLLSFSGRILQSEKI--HQGGKTFDYNPQFADWSKETRGAPLISVKP 493
               : |||| :  | |:    |||:  : |  | |       |  ||| :  |   :::
Db        416 VKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTPS 472

Qy        494 --LDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKK-AIMIEVDDRTEAYLRALQQKV 550
                || | :|  :||     :|: :|::    ||:::      |  |||| |:||:   |
Db        473 DGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCV 532

Qy        551 TSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMN 610
               || ::::||: ::  ::| :|||    |   |:| |  :|  | :::|:||||||:|:|
Db        533 RSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQLN 591

Qy        611 CKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASIN-EGMTRWFSRCVFQDRG 669
              ||:|   | ::: |   | :| |    |   :|:    :||:: :  : :||
Db        592 CKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAF 651

Qy        670 QELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVG 729
                | : |   :  ||| :   :  :|||:: ||||  | || |  :||   :: ||:
Db        652 DVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEY 711

Qy        730 RGYN---PRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVR 786
                    |:|  ||| : :| |||       ||| |||::|  :| || |||::||| ||
Db        712 ARVQLSPPQLAYIVVTRSMNTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVSQQVR 767

Qy        787 SGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVGQS 846
              |:|||| |||:| | |  |: :|:|||| ||:|||| |  |||| |||| ||| ||| :
Db        768 QGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTN 827

Qy        847 IHREPNLSLSNRLYYL 862
               :|  |  :|  : |||
Db        828 LHSIPQNALEKKFYYL 843


RESULT 6
AAU07866
ID   AAU07866 standard; Protein; 580 AA.
XX
AC   AAU07866;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Polypeptide sequence for mammalian Spg16.
XX
KW   Mammalian; reproductive-specific protein; male infertility;
KW   spermatogenesis; sperm count disorder; anti infertility; reproduction.
XX
OS   Mammalia.
XX
PN   WO200166752-A2.
XX
PD   13-SEP-2001.
XX
PF   07-MAR-2001; 2001WO-US07371.
XX
PR   07-MAR-2000; 2000US-0187518.
PR   12-JAN-2001; 2001US-0261557.
```

```
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX
PI   Wang PJ,  Page DC;
XX
DR   WPI; 2001-570774/64.
DR   N-PSDB; AAS13630.
XX
PT   Novel reproduction-specific protein, useful for treating disorders of
PT   reduced sperm count, enhancing/increasing sperm count and/or sperm
PT   activity -
XX
PS   Claim 22; Fig 16; 151pp; English.
XX
CC   The present invention relates to the isolation of novel mammalian and
CC   human reproductive-specific proteins (AAU07859-AAU07899), and the
CC   nucleic acids encoding them. The nucleic acids encoding
CC   reproductive-specific proteins are useful for diagnosing infertility
CC   which is a result of reduced sperm count, reduced sperm motility,
CC   malformed sperm or combinations of these. The sequences of the invention
CC   are useful as markers for spermatogonial cells, for identifying genes or
CC   proteins characteristic of male infertility, diagnosing or aiding in
CC   the diagnosis of infertility in men, and for contraception in which
CC   sperm production or sperm count is reduced or defective sperm is
CC   produced. Antibodies to reproductive-specific proteins are useful for
CC   determining the presence of these proteins in a sample obtained from a
CC   man being assessed for infertility, for identifying the expression of
CC   genes in particular cell type or particular developmental stage, for
CC   studies of spermatogenesis, and for immunofluorescence of germ cells or
CC   in Western blots for assessing the presence of the protein the antibody
CC   binds. The sequences of the invention are also useful for treating
CC   disorders of reduced sperm count, and for increasing sperm count and/or
CC   sperm activity. The nucleic acids of the invention are useful in gene
CC   therapy. AAU07859-AAU07882 represent the mammalian reproduction-specific
CC   proteins of the present invention.
XX
SQ   Sequence   580 AA;

  Query Match             30.0%;  Score 1360.5;  DB 22;  Length 580;
  Best Local Similarity   43.3%;  Pred. No. 1.7e-126;
  Matches  253;  Conservative  121;  Mismatches  201;  Indels    9;  Gaps    8;

Qy        284 MFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSF 343
              |  :||| :|| ||:: || |:| ||:|:|||:|||:||:||:: || :|  :|| |::|
Db          1 MHAIYQQNKEH-FQDECSKLLVGSIVITRYNNRTYRIDDVDWNKTPKDSFVMSDGKEITF  59

Qy        344 LEYYRKQYNQEITDLKQPVLVSQP-KRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFN 402
              |||| | |   : :  ||:|: :| :|:   |  | |  : :||| ::||: :||: ||
Db         60 LEYYSKNYGITVKEDDQPLLIHRPSERQNNHGMLLKGEILLLPELSFMTGIPEKMKKDFR 119

Qy        403 VMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEK 462
               ||||     |:|:|    :  |:  | :::    ||  ||||   ::    ||:|  |:
Db        120 AMKDLTQQINLSPKQHHGALECLLQRISQNETASNELTRWGLSLHKDVHKIEGRLLPMER 179

Qy        463 IHQGGKTFDYNPQFADWSKE-TRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTP 521
              |:    :|    :  :| || || | :::: |:  | | | :|  : |  |:  | |:
Db        180 INLRNTSF-VTSEGLNWVKEVTRDASILTI-PMHFWALFYPKRAMDQARELVNMLEKIAG 237

Qy        522 AMGIQMKKAIMIEV-DDRTEAYLRALQQ--KVTSDTQIVVCLLSSNRKDKYDAIKKYLCT 578
               :|::      :|: ||| | |:| :|    |    |:|||::   | | | ||||  |
Db        238 PIGMRTSPPAWVELKDDRIETYIRTIQSLLGVEGKIQMVVCIIMGTRDDLYGAIKKLCCV 297

Qy        579 DCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDT 638
               | ||| :  ||:|:   : ::| || ||||||:||||| ||: ||  |::|:| |||
Db        298 QSPVPSQVINVRTIGQPTRLRSVAQKILLQMNCKLGGELWGVDIPLKQLMVIGMDVYHDP 357

Qy        639 TAGRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRV 698
              : | ||: ||||||| :|:|:|| ||| ||:|| ||:||  :|: :  |  :| ::
Db        358 SRGMRSVVGFVASINLTLTKWYSRVVFQMPHQEIVDSLKLCLVGSLKKYYEVNHCLPEKI 417

Qy        699 IVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQ 758
              :|||||| |||||: |||:||   | ::    |:|:: | ||:|:::   :  :
Db        418 VVYRDGVSDGQLKTVANYEIPQLQKCFEAFD-NYHPKMVVFVVQKKISTNLYLAAPDHFV 476

Qy        759 NPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCH 818
              | ||||:|  :|  || ||:::  || |  |||| : ::: | |||:||||:| ||
Db        477 TPSPGTVVDHTITSCEWVDFYLLAHHVRQGCGIPTHYICVLNTANLSPDHMQRLTFKLCH 536

Qy        819 VYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 862
              :|:|||| |||||||:|||||||| || :| || : |   |::|
Db        537 MYWNWPGTIRVPAPCKYAHKLAFLSGQILHHEPAIQLCGNLFFL 580


RESULT 7
AAB94209
ID   AAB94209 standard; Protein; 530 AA.
XX
AC   AAB94209;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:14559.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saitó K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 14559; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
```

SQ Sequence 530 AA;

Query Match 26.3%; Score 1190; DB 22; Length 530;
Best Local Similarity 42.2%; Pred. No. 1.7e-109;
Matches 224; Conservative 103; Mismatches 198; Indels 6; Gaps 5;

```
Qy        336 ADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRR-GPGGTLPGPAMXIPELCYLTGLT 394
Db          2 SDGKEITFLEYYSKNYGITVKEEDQPLLIHRPSERQDNHGMLLKGEILLLPELSFMTGIP 61

Qy        395 DKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFS 454
Db         62 EKMKKDFRAMKDLAQQINLSPKQHHSALECLLQRIAKNEAATNELMRWGLRLQKDVHKIE 121

Qy        455 GRILQSEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQ 514
Db        122 GRVLPMERINLKNTSF-ITSQELNWVKEVTRDPSILTIPMHFWALFYPKRAMDQARELVN 180

Qy        515 NLFKVTPAMGIQMKKAIMIEV-DDRTEAYLRALQQKVTSD--TQIVVCLLSSNRKDKYDA 571
Db        181 MLEKIAGPIGMRMSPPAWVELKDDRIETYVRTIQSTLGAEGKIQMVVCIIMGPRDDLYGA 240

Qy        572 IKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMALKLAMIVG 631
Db        241 IKKLCCVQSPVPSQVVNVRTIGQPTRLRSVAQKILLQINCKLGGELWGVDIPLKQLMVIG 300

Qy        632 IDCYHDTTAGRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCN 691
Db        301 MDVYHDPSRGMRSVVGFVASINLTLTKWYSRVVFQMPHQEIVDSLKLCLVGSLKKFYEVN 360

Qy        692 EYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFA 751
Db        361 HCLPEKIVVYRDGVSDGQLKTVANYEIPQLQKCFEAF-ENYQPKMVVFVVQKKISTNLYL 419

Qy        752 QSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQR 811
Db        420 AAPQNFVTPTPGTVVDHTITSCEWVDFYLLAHHVRQGCGIPTHYVCVLNTANLSPDHMQR 479

Qy        812 LTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 862
Db        480 LTFKLCHMYWNWPGTIRVPAPCKYAHKLAFLSGHILHHGPAIQLCENLFFL 530
```

RESULT 8
AAB92722
ID AAB92722 standard; Protein; 371 AA.
XX
AC AAB92722;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:11144.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 11144; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 371 AA;

Query Match 20.6%; Score 932; DB 22; Length 371;
Best Local Similarity 45.5%; Pred. No. 6.3e-84;
Matches 168; Conservative 73; Mismatches 124; Indels 4; Gaps 3;

```
Qy        497 WLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEV-DDRTEAYLRALQQKVTSD-- 553
Db          4 WALFYPKRAMDQARELVNMLEKIAGPIGMRMSPPAWVELKDDRIETYVRTIQSTLGAEGK 63

Qy        554 TQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKM 613
Db         64 IQMVVCIIMGPRDDLYGAINKLCCVQSPVPSQVVNVRTIGQPTRLRSVAQKILLQINCKL 123

Qy        614 GGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWFSRCVFQDRGQELV 673
Db        124 GGELWGVDIPLKQLMVIGMDVYHDPSRGMRSVVGFVASINLTLTKWYSRVVFQMPHQEIV 183

Qy        674 DGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYN 733
Db        184 DSLKLCLVGSLKKFYEVNHCLPEKIVVYRDGVSDGQLKTVANYEIPQLQKCFEAF-ENYQ 242

Qy        734 PRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPT 793
Db        243 PKMVVFVVQKKISTNLYLAAPQNFVTPTPGTVVDHTITSCEWVDFYLLAHHVRQGCGIPT 302

Qy        794 HYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNL 853
Db        303 HYVCVLNTANLSPDHMQRLTFKLCHMYWNWPGTIRVPAPCKYAHKLAFLSGHILHHEPAI 362

Qy        854 SLSNRLYYL 862
Db        363 QLCENLFFL 371
```

RESULT 9
AAB58977
ID AAB58977 standard; Protein; 130 AA.
XX
AC AAB58977;

```
XX
DT   27-MAR-2001  (first entry)
XX
DE   Breast and ovarian cancer associated antigen protein sequence SEQ ID 685.
XX
KW   Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW   nootropic; neurpprotective; antiviral; antiallergic; hepatotropic;
KW   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW   antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW   Addison's disease; allergy; autoimmune haemolytic anaemia;
KW   autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW   multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW   cardiovascular disorder; wound healing; neurological disease.
XX
OS   Homo sapiens.
XX
PN   WO200055173-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05881.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-611515/58.
DR   N-PSDB; AAF21880.
XX
PT   New human breast and ovarian cancer associated gene sequences and the
PT   polypeptides encoded by these genes, useful in the prevention,
PT   treatment and diagnosis of cancer, immune disorders, cardiovascular
PT   disorders and neurological diseases -
XX
PS   Claim 11; Page 1139; 1299pp; English.
XX
CC   Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC   proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC   associated with breast and ovarian cancer. Included in the invention are
CC   sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC   isolation and characterisation of the DNA and protein sequences of the
CC   invention. The breast and ovarian cancer associated DNA, protein, agonist
CC   or antagonist sequences exhibit cytostatic; immunosuppressive;
CC   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
CC   antibacterial; antifungal; antiparasitic and cardiant activity. The
CC   polynucleotide and protein sequences are used in the diagnosis of cancer,
CC   particularly breast and ovarian cancer. The nucleic acid sequences,
CC   proteins, agonists and agonists may also be used in the diagnosis,
CC   prevention and treatment of immune disorders e.g. Addison's disease,
CC   allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC   diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC   arthritis and ulcerative colitis; cardiovascular disorders such as
CC   myocardial ischaemias; wound healing; neurological diseases such as
CC   cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ   Sequence   130 AA;

  Query Match             13.7%;  Score 620;  DB 21;  Length 130;
  Best Local Similarity   92.7%;  Pred. No. 1.7e-53;
  Matches  115;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        579 DCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDT 638
              ||||||||||||||||||||||||||||||||||||||||||: ||| ||||||| ||
Db          5 DCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCXHDM 64

Qy        639 TAGRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRV 698
              ||||||||||||||||||||||:||||||||||||||||||: ||||||||:
Db         65 TAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRI 124

Qy        699 IVYR 702
              ||||
Db        125 IVYR 128


RESULT 10
ABG05957
ID   ABG05957 standard; Protein; 341 AA.
XX
AC   ABG05957;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #5948.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS70144.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 36316; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   341 AA;

  Query Match             12.8%;  Score 579;  DB 22;  Length 341;
  Best Local Similarity   43.9%;  Pred. No. 1.2e-48;
  Matches  115;  Conservative   49;  Mismatches   64;  Indels   34;  Gaps    3;

Qy        176 QHKVTEVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYN 235
              |  |||: |:|: || :::|||| ||| :|| |:| :|| ||
Db         16 QPTVTELSSETQRGETIKMTITLKRELPSSSPVCIQVFNIIFR---------------- 58
```

```
Qy        236 PSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHK 295
                        :| :|||| |: :| :: |||:||||:||||:|| | |:|
Db         59 ----------KLSLWPGFAISVSYFERKLLFSADVSYKVLRNETVLEFMTALCQRTGLSC 108

Qy        296 FQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEI 355
              | :  |:||||||||:|||:|| :|||||   |  ||:| ||:|:::::||::||:  :
Db        109 FTQTCEKQLIGLIVLTRYNNRTYSIDDIDWSVKPTHTFQKRDGTEITYVDYYKQQYDITV 168

Qy        356 TDLKQPVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTP 415
              :|| ||:||| |::| |  |||||:|||||:  :|| :|| :|  |||:|
Db        169 SDLNQPMLVSLLKKKRNDNSEAQ-LAHLIPELCFLTGLTDQATSDFQLMKAVAEKTRLSP 227

Qy        416 EQRQREVGRLIDYIHKDDNVQR 437
                ||: : ||:      ||:||
Db        228 SGRQQRLARLV------DNIQR 243
```

RESULT 11
AAG51562
ID AAG51562 standard; Protein; 983 AA.
XX
AC AAG51562;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 65455.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.

PR 29-OCT-1999; 99US-0162142.

Query Match 11.2%; Score 506.5; DB 21; Length 983;
Best Local Similarity 23.6%; Pred. No. 1.4e-40;
Matches 224; Conservative 144; Mismatches 426; Indels 155; Gaps 36;

```
Qy          4 RARARARGRARGQETVQ-HVGAAASQQPGYIPPRPQQSPTEGDLVGRGRQRGMVVGATSK 62
              |:| | ||   |::: |  |    | :|   || | |:                  ||
Db         60 RSRRRNRG---GRKSDQGDVCMRPSSRPRK-PPPPSQT-----------------TSS 96

Qy         63 SQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHFRL 122
              :          :|:|  |       |   :  |:| :        |: |   :  ||| |
Db         97 A----------VSVATAGEIVAVNHQMQMGVRKNSNFAPRPGFGTLGTKCIVKANHF-L 144

Qy        123 TSRPQWALYQYHIDYNPLMEARRLRSAL------LFQHEDLIGRCHAFDGT--------- 167
                |   | || :   | : :: :  |:      |::  ||  |  |:||
Db        145 ADLPTKDLNQYDVTITPEVSSKSVNRAIIAELVRLYKESDLGRRLPAYDGRKSLYTAGEL 204

Qy        168 ----------------ILFLPKR-LQHKVTEVFSQTRNGEHVRITITLTNELPPTSPTC 209
                             |:  |||   :||   |    |  |  :   |  :
Db        205 PFTWKEFSVKIVDEDDGIINGPKRERSYKVAIKFVARANMHH--LGEFLAGKRADCPQEA 262

Qy        210 LQFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIP---NHRLVIWPGFTTSILQYENNIML 266
              :|  :|  |  | :     :||::::|   |  |       |  | ||  ||  :  : |
Db        263 VQILDIVLRE-LSVKRFCPVGRSFFSPD--IKTPQRLGEGLESWCGFYQSIRPTQMGLSL 319

Qy        267 CTDV-SHKVLRSETVLDFMFNLYQQTEEHK-----FQEQVSKELIGLIVLTKYN---NKT 317
                |: |   :    |::|:  |  :    |      : :: | | |: |   :.    :
Db        320 NIDMASAAFIEPLPVIEFVAQLLGKDVLSKPLSDSDRVKIKKGLRGVKVEVTHRANVRRK 379

Qy        318 YRVDDIDWDQNPKSTFKKADGSEV-SFLEYYRKQYNQEITDLKQPVL-VSQPKRRRGPGG 375
              |||  :      :  |   :   : | :||::: |   |     | | |   |:
Db        380 YRVAGLTTQPTRELMFPVDENCTMKSVIEYFQEMYGFTIQHTHLPCLQVGNQKK----AS 435

Qy        376 TLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNV 435
              ||  |  | |    |    |  |:  :   | | |   |  |: :: | : :   | :
Db        436 YLPMEACKIVEGQRYT----KRLNEKQITALLKV-TCQRPRDRENDILRTVQHNAYDQDP 490

Qy        436 QRELRDWGLSFDSNLLSFSGRILQSE--KIHQGGKTFDYNPQFADWSKETRGAPLISVKP 493
                :::|::    | |   ||| :   | |: ||  |  ||   |:   :   :|:
Db        491 Y--AKEFGMNISEKLASVEARILPAPWLKYHENGKEKDCLPQVGQWNMMNK--KMINGMT 546

Qy        494 LDNWLLIYTRRNYE--AANSLIQNLFKVTPAMGIQMKKAIMIEV----DDRTE-----AY 542
              :  |  :   |: :   |      | ::    |::    :| :     |: |      |
Db        547 VSRWACVNFSRSVQENVARGFCNELGQMCEVSGMEFNPEPVIPIYSARPDQVEKALKHVY 606

Qy        543 LRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVAR---TLGKQQTVM 599
              :: :    : :::: :|  |    |  :|:   |:    |||  : :    : ||
Db        607 HTSMNKTKGKELELLLAILPDNNGSLYGDLKRICETELGLISQCCLTKHVFKISKQ---- 662

Qy        600 AIATKIALQMNCKMGGE--------LWRVDMALKLAMIV-GIDCYHDTTA--GRRSIAGF 648
                  ::|::| ||||           |: :   :  |: | |  |          |||
Db        663 -YLANVSLKINVKMGGRNTVLVDAISCRIPLVSDIPTIIFGADVTHPENGEESSPSIAAV 721

Qy        649 VASIN-EGMTRWFSRCVFQDRGQELVDGLKVCLQAALR-------------AWSGCNEYM 694
              ||| : . :|::       |   |||:  |    |  :|             ::
Db        722 VASQDWPEVTKYAGLVCAQAHRQELIQDLYKTWQDPVRGTVSGGMIRDLLISFRKATGQK 781

Qy        695 PSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFA--- 751
              | |:| ||||| :||   :: ||:         |:   | | :| |||:|| : | ||
Db        782 PLRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEPNYQPPVTFIVVQKRHHTRLFANNH 841

Qy        752 ---QSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDH 808
                  |  |  | |||||:| ::  |  :||:: | |   |:  | ||:|::| :    |
Db        842 RDKNSTDRSGNILPGTVVDTKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDENNFTADG 901

Qy        809 IQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSN 857
              || ||   |: |      : :  |  |||  ||      : || :   |
Db        902 IQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRA--RFYLEPEIMQDN 948
```

RESULT 12
AAG51561

ID AAG51561 standard; Protein; 988 AA.
XX
AC AAG51561;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 65454.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.

```
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match           11.2%;  Score 506.5;  DB 21;  Length 988;
  Best Local Similarity  23.6%;  Pred. No. 1.4e-40;
  Matches  224;  Conservative  144;  Mismatches  426;  Indels  155;  Gaps   36;

Qy          4 RARARARGRARGQETVQ-HVGAAASQQPGYIPPRPQQSPTEGDLVGRGRQRGMVVGATSK 62
              |:| | ||   |::: |  |    | :|   || | |:                    ||
Db         65 RSRRRNRG---GRKSDQGDVCMRPSSRPRK-PPPPSQT------------------TSS 101

Qy         63 SQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHFRL 122
              :           :|:|  |      |   :  |:| :       |: |    :  ||| |
Db        102 A-----------VSVATAGEIVAVNHQMQMGVRKNSNFAPRPGFGTLGTKCIVKANHF-L 149

Qy        123 TSRPQWALYQYHIDYNPLMEARRLRSAL------LFQHEDLIGRCHAFDGT--------- 167
                |   | ||  :   | : :: :  |:      |::  ||  |  |:||
Db        150 ADLPTKDLNQYDVTITPEVSSKSVNRAIIAELVRLYKESDLGRRLPAYDGRKSLYTAGEL 209

Qy        168 -----------------ILFLPKR-LQHKVTEVFSQTRNGEHVRITITLTNELPPTSPTC 209
                             |:  |||   :||   |    |  |  :   |  :

Db        210 PFTWKEFSVKIVDEDDGIINGPKRERSYKVAIKFVARANMHH--LGEFLAGKRADCPQEA 267

Qy        210 LQFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIP---NHRLVIWPGFTTSILQYENNIML 266
              :|  :|  |  | :     :||::::|   |  |       |  | ||  ||   :  : |
Db        268 VQILDIVLRE-LSVKRFCPVGRSFFSPD--IKTPQRLGEGLESWCGFYQSIRPTQMGLSL 324

Qy        267 CTDV-SHKVLRSETVLDFMFNLYQQTEEHK-----FQEQVSKELIGLIVLTKYN---NKT 317
              |: |   :    |::|:  |  :   |       : :: | | |: |   :     :
Db        325 NIDMASAAFIEPLPVIEFVAQLLGKDVLSKPLSDSDRVKIKKGLRGVKVEVTHRANVRRK 384

Qy        318 YRVDDIDWDQNPKSTFKKADGSEV-SFLEYYRKQYNQEITDLKQPVL-VSQPKRRRGPGG 375
              |||  :      :  |   :   : | :||::: |   |      | | |   |:
Db        385 YRVAGLTTQPTRELMFPVDENCTMKSVIEYFQEMYGFTIQHTHLPCLQVGNQKK----AS 440

Qy        376 TLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNV 435
              ||  |  | |    |    |  |:  :   | | |   |  |: :: | : :   | :
Db        441 YLPMEACKIVEGQRYT----KRLNEKQITALLKV-TCQRPRDRENDILRTVQHNAYDQDP 495

Qy        436 QRELRDWGLSFDSNLLSFSGRILQSE--KIHQGGKTFDYNPQFADWSKETRGAPLISVKP 493
                :::|::    | |   ||| :   | |: ||  |  ||   |:   :  :|:
Db        496 Y--AKEFGMNISEKLASVEARILPAPWLKYHENGKEKDCLPQVGQWNMMNK--KMINGMT 551

Qy        494 LDNWLLIYTRRNYE--AANSLIQNLFKVTPAMGIQMKKAIMIEV----DDRTE-----AY 542
              :  |  :   |: :   |     | ::   |::    :| :      |: |      |
Db        552 VSRWACVNFSRSVQENVARGFCNELGQMCEVSGMEFNPEPVIPIYSARPDQVEKALKHVY 611

Qy        543 LRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVAR---TLGKQQTVM 599
              :: :    : :::: :|  |    |  :|:   |:    ||| : :    : ||
Db        612 HTSMNKTKGKELELLLAILPDNNGSLYGDLKRICETELGLISQCCLTKHVFKISKQ---- 667

Qy        600 AIATKIALQMNCKMGGE--------LWRVDMALKLAMIV-GIDCYHDTTA--GRRSIAGF 648
                  ::|::| ||||           |: :   :  |: | |  |          |||
Db        668 -YLANVSLKINVKMGGRNTVLVDAISCRIPLVSDIPTIIFGADVTHPENGEESSPSIAAV 726

Qy        649 VASIN-EGMTRWFSRCVFQDRGQELVDGLKVCLQAALR------------AWSGCNEYM 694
              ||| :   :|::       |   |||:  |    | :|             ::
Db        727 VASQDWPEVTKYAGLVCAQAHRQELIQDLYKTWQDPVRGTVSGGMIRDLLISFRKATGQK 786

Qy        695 PSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFA--- 751
              | |:| ||||| :||   :: ||:         |:   | | :| |||:|| : | ||
Db        787 PLRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEPNYQPPVTFIVVQKRHHTRLFANNH 846

Qy        752 ---QSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDH 808
                |  |  | |||||:| ::  |  :||:: | |    |:  | ||:|::| :    |
Db        847 RDKNSTDRSGNILPGTVVDTKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDENNFTADG 906

Qy        809 IQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSN 857
              || ||   |: |      : :  |  |||  ||      : || :   |
Db        907 IQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRA--RFYLEPEIMQDN 953


RESULT 13
AAG51563
ID   AAG51563 standard; Protein; 906 AA.
XX
AC   AAG51563;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 65456.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
```

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

```
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match           11.0%;  Score 500;  DB 21;  Length 906;
  Best Local Similarity  23.7%;  Pred. No. 5.3e-40;
  Matches  205;  Conservative  135;  Mismatches  404;  Indels  120;  Gaps   31;

Qy        88 HDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLR 147
             |   :  |:| :       |: |    :  ||| |   |  | || :   | : :: :
Db        34 HQMQMGVRKNSNFAPRPGFGTLGTKCIVKANHF-LADLPTKDLNQYDVTITPEVSSKSVN 92

Qy       148 SAL------LFQHEDLIGRCHAFDGT-------------------------ILFLPKR- 174
              |:      |::  ||  |  |:||                          |:  |||
Db        93 RAIIAELVRLYKESDLGRRLPAYDGRKSLYTAGELPFTWKEFSVKIVDEDDGIINGPKRE 152

Qy       175 LQHKVTEVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYY 234
             :||   |    |  |  :   |  :       :|  :|  |  | :     :||:::
Db       153 RSYKVAIKFVARANMHH--LGEFLAGKRADCPQEAVQILDIVLRE-LSVKRFCPVGRSFF 209

Qy       235 NPSDPIDIP---NHRLVIWPGFTTSILQYENNIMLCTDV-SHKVLRSETVLDFMFNLYQQ 290
             :|   |  |      |  | ||  ||   :  : |  |: |   :    |::|:  |  :
Db       210 SPD--IKTPQRLGEGLESWCGFYQSIRPTQMGLSLNIDMASAAFIEPLPVIEFVAQLLGK 267

Qy       291 TEEHK-----FQEQVSKELIGLIVLTKYN---NKTYRVDDIDWDQNPKSTFKKADGSEV- 341
               |      : :: | | |: |   :     : |||  :      :  |   :   :
Db       268 DVLSKPLSDSDRVKIKKGLRGVKVEVTHRANVRRKYRVAGLTTQPTRELMFPVDENCTMK 327

Qy       342 SFLEYYRKQYNQEITDLKQPVL-VSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRND 400
             | :||::: |   |     | | |   |:       ||  |  | |    |    |  |:
Db       328 SVIEYFQEMYGFTIQHTHLPCLQVGNQKK----ASYLPMEACKIVEGQRYT----KRLNE 379

Qy       401 FNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQS 460
               :  | | |   |  |: :: | : :   | :     :::|::    | |   ||| :
Db       380 KQITALLKV-TCQRPRDRENDILRTVQHNAYDQDPY--AKEFGMNISEKLASVEARILPA 436

Qy       461 E--KIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYE--AANSLIQNL 516
               | |: ||  |  ||   |:   :  :|:   :  | :   |: :   |      |
Db       437 PWLKYHENGKEKDCLPQVGQWNMMNK--KMINGMTVSRWACVNFSRSVQENVARGFCNEL 494

Qy       517 FKVTPAMGIQMKKAIMIEV----DDRTE-----AYLRALQQKVTSDTQIVVCLLSSNRKD 567
              ::    |::     :| :     |: |      |  :: :    : :::: :|  |

Db       495 GQMCEVSGMEFNPEPVIPIYSARPDQVEKALKHVYHTSMNKTKGKELELLLAILPDNNGS 554

Qy       568 KYDAIKKYLCTDCPTPSQCVVAR---TLGKQQTVMAIATKIALQMNCKMGGE-------- 616
               |  :|:   |:    ||| : :    : ||         ::|::| ||||
Db       555 LYGDLKRICETELGLISQCCLTKHVFKISKQ-----YLANVSLKINVKMGGRNTVLVDAI 609

Qy       617 LWRVDMALKLAMIV-GIDCYHDTTA--GRRSIAGFVASIN-EGMTRWFSRCVFQDRGQEL 672
               |: :   :  |: | |  |          |||  ||| :   :|::      |   |||
Db       610 SCRIPLVSDIPTIIFGADVTHPENGEESSPSIAAVVASQDWPEVTKYAGLVCAQAHRQEL 669

Qy       673 VDGLKVCLQAALR------------AWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVP 719
             :  |    |  :|            ::        | |:| ||||| :||   :: ||:
Db       670 IQDLYKTWQDPVRGTVSGGMIRDLLISFRKATGQKPLRIIFYRDGVSEGQFYQVLLYELD 729

Qy       720 QFLDCLKSVGRGYNPRLTVIVVKKRVNARFFA------QSGGRLQNPLPGTVIDVEVTRP 773
                   |:   | | :| |||:|| : | ||       |  |  | |||||:| ::  |
Db       730 AIRKACASLEPNYQPPVTFIVVQKRHHTRLFANNHRDKNSTDRSGNILPGTVVDTKICHP 789

Qy       774 EWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPC 833
               :||:: | |    |:  | ||:|::|  :    | || ||   |: |      : :  |
Db       790 TEFDFYLCSHAGIQGTSRPAHYHVLWDENNFTADGIQSLTNNLCYTYARCTRSVSIVPPA 849

Qy       834 QYAHKLAFLVGQSIHREPNLSLSN 857
               |||  ||      : || :   |
Db       850 YYAHLAAFRA--RFYLEPEIMQDN 871


RESULT 14
AAG42118
ID   AAG42118 standard; Protein; 1048 AA.
XX
AC   AAG42118;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 52487.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
```

PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.

```
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match           11.0%;  Score 499.5;  DB 21;  Length 1048;
  Best Local Similarity   23.5%;  Pred. No. 7.6e-40;
  Matches  240;  Conservative  156;  Mismatches  425;  Indels  199;  Gaps   42;

Qy          3 GRARARARGRARG---QETVQHVGAAA-SQQP--------GY-IPPRPQQS--------- 40
              |      ||  ||   |   || |    : ||         ||  ||: ||
Db         23 GPVSGGGRGSQRGGFQQGGGQHQGGRGYTPQPQQGGRGGRGYGQPPQQQQQYGGPQEYQG 82

Qy         41 ------PTEGDLVGRGRQRG--------------------------------MVVGATS 61
                  | :|   | |  ||                                  |  |
Db         83 RGRGGPPHQGGRGGYGGGRGGGPSSGPPQRQSVPELHQATSPTYQAVSSQPTLSEVSPTQ 142

Qy         62 KSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHFR 121
               :   ::  |::||: |:|   :       :::     ::  | | ||    :  |||
Db        143 VPEPTVLAQQFEQLSV-EQGAPSQAIQPIPSSSKAFKFPMRPGK-GQSGKRCIVKANHF- 199

Qy        122 LTSRPQWALYQYHIDYNPLMEARRLRSALLFQHED------LIGRCHAFDG-----TILF 170
                  |   |: | :   | : :| :  |:: |  |       |  |  |:||     |
Db        200 FAELPDKDLHHYDVTITPEVTSRGVNRAVMKQLVDNYRDSHLGSRLPAYDGRKSLYTAGP 259

Qy        171 LP------------------KRLQHKVTEVFSQTRNGEHVRITITLTNELPPTSPTCLQ 211
              ||                 :| : :   |       :   : : |  :         ||
Db        260 LPFNSKEFRINLLDEEVGAGGQRREREFKVVIKLVARADLHHLGMFLEGKQSDAPQEALQ 319

Qy        212 FYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIPNHR-----LVIWPGFTTSILQYENNIML 266
               :|  |  |       :||::|:|    ||   :     |  | ||  ||   :  : |
Db        320 VLDIVLRE-LPTSRYIPVGRSFYSP----DIGKKQSLGDGLESWRGFYQSIRPTQMGLSL 374

Qy        267 CTDVSHKV-LRSETVLDFMFNLYQQTEEHK-----FQEQVSKELIGL-IVLTKYNN--KT 317
                |:|    : :  |: |: :|  :    :       : :: | | |: : :|   |  :
Db        375 NIDMSSTAFIEANPVIQFVCDLLNRDISSRPLSDADRVKIKKALRGVKVEVTHRGNMRRK 434

Qy        318 YRVDDIDWDQNPKSTFKKAD-GSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGT 376
              ||:  :      : ||    :  :: | :||: : |   |   : | |       |
Db        435 YRISGLTAVATRELTFPVDERNTQKSVVEYFHETYGFRIQHTQLPCLQVGNSNR------ 488

Qy        377 LPGPAMXIPELCYLT--GLTDKMRNDFNVMKDLAVHTRLTPEQRQREV---GRLIDYIHK 431
               |     |:| :       |  |:  :   | | |   |  |::::    :| ||  |
Db        489 ---PNYLPMEVCKIVEGQRYSKRLNERQITALLKV-TCQRPIDREKDILQTVQLNDYA-K 543

Qy        432 DDNVQRELRDWGLSFDSNLLSFSGRILQSE--KIHQGGKTFDYNPQFADWSKETRGAPLI 489
              |:  |    ::|:   ::| |   |||     | |: |:     ||   |:   :   :|
Db        544 DNYAQ----EFGIKISTSLASVEARILPPPWLKYHESGREGTCLPQVGQWNMMNK--KMI 597

Qy        490 SVKPLDNWLLIYTRRNYE--AANSLIQNLFKVTPAMGIQMK-KAIMIEVDDRTEAYLRAL 546
              :   ::||: |   |  :   | :  | | ::     |:    : ::  |  | |   : |
Db        598 NGGTVNNWICINFSRQVQDNLARTFCQELAQMCYVSGMAFNPEPVLPPVSARPEQVEKVL 657

Qy        547 QQK---VTS------DTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGK-QQ 596
              :  :    ||     :  ::: :|  |     |  :|:    |:  ·  ||| : : |  :
Db        658 KTRYHDATSKLSQGKEIDLLIVILPDNNGSLYGDLKRICETELGIVSQCCLTKHVFKMSK 717

Qy        597 TVMAIATKIALQMNCKMGGELWRVDMAL---------KLAMIVGIDCYH-----DTTAGR 642
                ||    :||::| |:||    :  ||          :  :| | |  |     |::
Db        718 QYMA---NVALKINVKVGGRNTVLVDALSRRIPLVSDRPTIIFGADVTHPHPGEDSSP-- 772

Qy        643 RSIAGFVASIN-EGMTRWFSRCVFQDRGQELVD----------------GLKVCLQAALR 685
               |||  |||  :   :|::       |   |||:                |:   |  | |
Db        773 -SIAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFKEWKDPQKGVVTGGMIKELLIAFR 831

Qy        686 AWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRV 745
               :|   : | |:| |||| :||   :: ||:         |:  || | :| :||:||
```

```
Db        832 RSTG---HKPLRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEAGYQPPVTFVVVQKRH 888

Qy        746 NARFFAQ------SGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIY 799
              : | |||      |  |  | ||||:| ::  |  :||:: | |   |:  | ||:|::
Db        889 HTRLFAQNHNDRHSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLW 948

Qy        800 DSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRL 859
              | :   | :| ||   |: |     :  |  |||  ||      : ||  | |  :
Db        949 DENNFTADGLQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRA--RFYMEPETSDSGSM 1006


RESULT 15
AAG52593
ID   AAG52593 standard; Protein; 1064 AA.
XX
AC   AAG52593;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 66875.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
```

PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 10.9%; Score 493.5; DB 21; Length 1064;
Best Local Similarity 23.1%; Pred. No. 3.1e-39;
Matches 240; Conservative 161; Mismatches 418; Indels 219; Gaps 43;

Qy 3 GRARARARGRARG---QETVQHVGAAA-SQQP--------GY-IPPRPQQS--------- 40

```
Db         23 GPVSGGGRGSQRGGFQQGGGQHQGGRGYTPQPQQGGRGGRGYGQPPQQQQQYGGPQEYQG 82

Qy         41 ------PTEGDLVGRGRQRG--------------------------------MVVGATS 61

Db         83 RGRGGPPHQGGRGGYGGGRGGGPSSGPPQRQSVPELHQATSPTYQAVSSQPTLSEVSPTQ 142

Qy         62 KSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHFR 121

Db        143 VPEPTVLAQQFEQLSV-EQGAPSQAIQPIPSSSKAFKFPMRPGK-GQSGKRCIVKANHF- 199

Qy        122 LTSRPQWALYQYHIDYNPLMEARRLRSALLFQHED------LIGRCHAFDG--------- 166

Db        200 FAELPDKDLHHYDVTITPEVTSRGVNRAVMKQLVDNYRDSHLGSRLPAYDGRKSLYTAGP 259

Qy        167 ----------------------------TILFLPKRLQHKVTEVFSQTRNG----EHVR 193

Db        260 LPFNSKEFRINLLDEEVGAGGQRREREFKVVIKLVARADLHHLGMFLEGKQSDAPQEALQ 319

Qy        194 ITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIPNHR-----LV 248

Db        320 VLDIVLRELPTSRYWLLILW--MFASLLPCLYI-PVGRSFYSP----DIGKKQSLGDGLE 372

Qy        249 IWPGFTTSILQYENNIMLCTDVSHKV-LRSETVLDFMFNLYQQTEEHK-----FQEQVSK 302

Db        373 SWRGFYQSIRPTQMGLSLNIDMSSTAFIEANPVIQFVCDLLNRDISSRPLSDADRVKIKK 432

Qy        303 ELIGL-IVLTKYNN--KTYRVDDIDWDQNPKSTFKKAD-GSEVSFLEYYRKQYNQEITDL 358

Db        433 ALRGVKVEVTHRGNMRRKYRISGLTAVATRELTFPVDERNTQKSVVEYFHETYGFRIQHT 492

Qy        359 KQPVLVSQPKRRRGPGGTLPGPAMXIPELCYLT--GLTDKMRNDFNVMKDLAVHTRLTPE 416

Db        493 QLPCLQVGNSNR---------PNYLPMEVCKIVEGQRYSKRLNERQITALLKV-TCQRPI 542

Qy        417 QRQREV---GRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSE--KIHQGGKTFD 471

Db        543 DREKDILQTVQLNDYA-KDNYAQ----EFGIKISTSLASVEARILPPPWLKYHESGREGT 597

Qy        472 YNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYE--AANSLIQNLFKVTPAMGIQMK- 528

Db        598 CLPQVGQWNMMNK--KMINGGTVNNWICINFSRQVQDNLARTFCQELAQMCYVSGMAFNP 655

Qy        529 KAIMIEVDDRTEAYLRALQQK---VTS------DTQIVVCLLSSNRKDKYDAIKKYLCTD 579

Db        656 EPVLPPVSARPEQVEKVLKTRYHDATSKLSQGKEIDLLIVILPDNNGSLYGDLKRICETE 715

Qy        580 CPTPSQCVVARTLGK-QQTVMAIATKIALQMNCKMGGELWRVDMAL---------KLAMI 629

Db        716 LGIVSQCCLTKHVFKMSKQYMA---NVALKINVKVGGRNTVLVDALSRRIPLVSDRPTII 772

Qy        630 VGIDCYH-----DTTAGRRSIAGFVASIN-EGMTRWFSRCVFQDRGQELVD--------- 674

Db        773 FGADVTHPHPGEDSSP---SIAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFKEWKDP 829

Qy        675 -------GLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKS 727

Db        830 QKGVVTGGMIKELLIAFRRSTG---HKPLRIIFYRDGVSEGQFYQVLLYELDAIRKACAS 886

Qy        728 VGRGYNPRLTVIVVKKRVNARFFAQ------SGGRLQNPLPGTVIDVEVTRPEWYDFFIV 781

Db        887 LEAGYQPPVTFVVVQKRHHTRLFAQNHNDRHSVDRSGNILPGTVVDSKICHPTEFDFYLC 946

Qy        782 SQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAF 841

Db        947 SHAGIQGTSRPAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAF 1006

Qy        842 LVGQSIHREPNLSLSNRL 859

Db       1007 RA--RFYMEPETSDSGSM 1022
```

Search completed: July 22, 2003, 07:01:03

Job time : 31.2338 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 06:59:33 ; Search time 11.4217 Seconds
(without alignments)
2220.565 Million cell updates/sec

Title: US-09-873-737A-4
Perfect score: 4528
Sequence: 1 MTGRARARARGRARGQETVQ.....VGQSIHREPNLSLSNRLYYL 862

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	131.5	2.9	3457	2	US-08-416-603-4	Sequence 4, Appli
2	110.5	2.4	560	2	US-08-756-317-11	Sequence 11, Appl
3	108	2.4	1253	1	US-08-252-966B-12	Sequence 12, Appl
4	108	2.4	1261	1	US-08-252-966B-18	Sequence 18, Appl
5	106	2.3	587	1	US-08-844-280-2	Sequence 2, Appli
6	106	2.3	587	4	US-09-006-726-2	Sequence 2, Appli
7	102.5	2.3	1180	3	US-09-224-024-28	Sequence 28, Appl
8	102.5	2.3	1180	5	PCT-US94-07902-28	Sequence 28, Appl
9	101.5	2.2	905	4	US-09-134-001C-3782	Sequence 3782, Ap
10	99	2.2	351	3	US-08-591-685-7	Sequence 7, Appli
11	98.5	2.2	731	1	US-08-731-716-2	Sequence 2, Appli
12	98.5	2.2	2285	4	US-09-308-375-2	Sequence 2, Appli
13	98	2.2	600	4	US-09-134-001C-3636	Sequence 3636, Ap
14	98	2.2	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
15	97.5	2.2	652	3	US-08-996-441B-6	Sequence 6, Appli
16	97.5	2.2	652	3	US-08-993-722A-6	Sequence 6, Appli
17	97.5	2.2	652	3	US-08-993-170A-6	Sequence 6, Appli
18	97.5	2.2	652	3	US-08-993-775B-6	Sequence 6, Appli
19	97.5	2.2	805	4	US-09-134-001C-4821	Sequence 4821, Ap
20	97	2.1	694	1	US-08-171-382-2	Sequence 2, Appli
21	97	2.1	694	1	US-08-309-420-2	Sequence 2, Appli
22	97	2.1	694	1	US-08-309-419-2	Sequence 2, Appli
23	97	2.1	694	4	US-09-294-531B-4	Sequence 4, Appli
24	97	2.1	694	5	PCT-US95-11856-2	Sequence 2, Appli
25	97	2.1	694	5	PCT-US95-11878-2	Sequence 2, Appli
26	96	2.1	1101	4	US-09-770-170-8	Sequence 8, Appli
27	95.5	2.1	615	4	US-09-134-001C-3284	Sequence 3284, Ap
28	95	2.1	928	5	PCT-US94-10357-3	Sequence 3, Appli
29	95	2.1	1151	4	US-09-134-001C-3242	Sequence 3242, Ap
30	94.5	2.1	648	1	US-08-451-715A-4	Sequence 4, Appli
31	94	2.1	1781	2	US-08-477-451-11	Sequence 11, Appl
32	93.5	2.1	1030	4	US-09-091-117-2	Sequence 2, Appli
33	93	2.1	749	1	US-08-276-099A-15	Sequence 15, Appl
34	93	2.1	749	1	US-08-781-890-15	Sequence 15, Appl
35	93	2.1	749	3	US-09-012-710-9	Sequence 9, Appli
36	93	2.1	749	4	US-09-364-970-4	Sequence 4, Appli
37	93	2.1	749	4	US-09-556-273-9	Sequence 9, Appli
38	93	2.1	749	4	US-08-956-653A-10	Sequence 10, Appl
39	92.5	2.0	321	3	US-08-673-814-2	Sequence 2, Appli
40	92.5	2.0	321	4	US-09-115-824-2	Sequence 2, Appli
41	92.5	2.0	457	3	US-09-057-969-5	Sequence 5, Appli
42	92.5	2.0	539	3	US-09-057-969-4	Sequence 4, Appli
43	92.5	2.0	624	3	US-09-057-969-3	Sequence 3, Appli
44	92.5	2.0	652	3	US-08-996-441B-32	Sequence 32, Appl
45	92.5	2.0	652	3	US-08-996-441B-34	Sequence 34, Appl

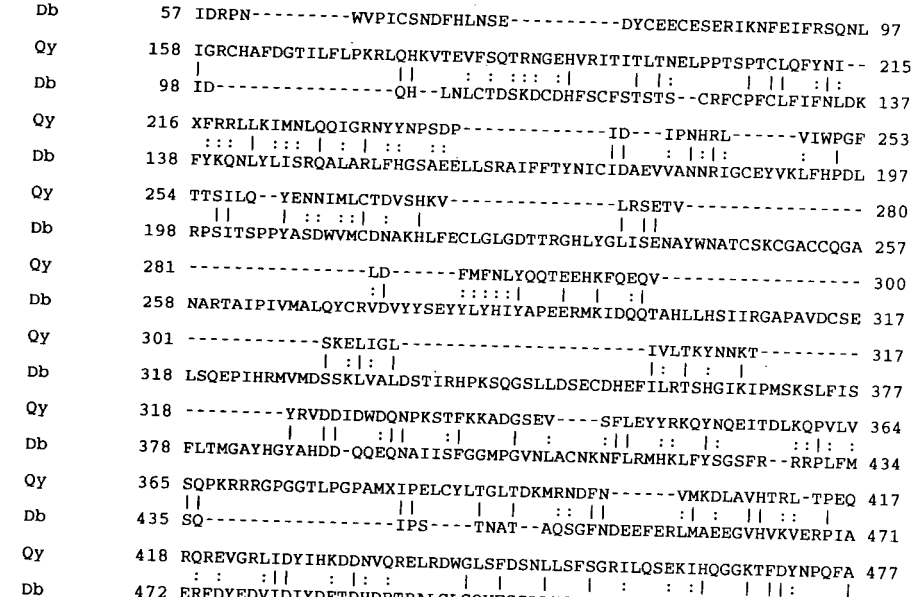
ALIGNMENTS

RESULT 1
US-08-416-603-4
; Sequence 4, Application US/08416603
; Patent No. 5866780
; GENERAL INFORMATION:
; APPLICANT: Law, Marcus
; APPLICANT: Hebara, Ledare
; APPLICANT: Reddick, Bradford B.
; TITLE OF INVENTION: Maize Chlorotic Dwarf Virus Genome and
; TITLE OF INVENTION: Uses Therefor
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/416,603
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lloyd, Jeffrey
; REGISTRATION NUMBER: 35,589
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3457 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-416-603-4

Query Match 2.9%; Score 131.5; DB 2; Length 3457;
Best Local Similarity 18.0%; Pred. No. 0.0056;
Matches 166; Conservative 129; Mismatches 283; Indels 343; Gaps 45;

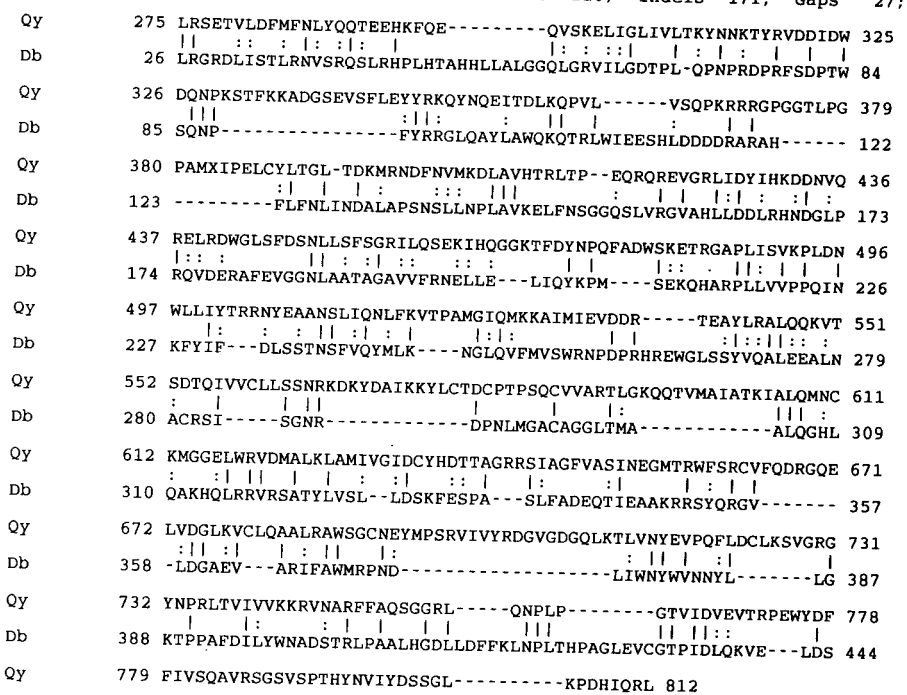
```
Qy          39 QSPTEGDLVGRGRQRGMVVGATSKSQELQISAGFQELSLAERGGRRRDFH-DXGVNTRQN 97
               |:||:| :              :|: : | : ||:  ||: :  | | |||      |:
Db           8 QNPTQGSIPEN----------SSQDRNLGVPAGY-SLSVEDPFGNRSDFHIPVHQIIREE 56

Qy          98 LDHVKESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDL 157
               :|            | : :| | | |          ||    |: |:::  :|: ::|
```



RESULT 2
US-08-756-317-11
; Sequence 11, Application US/08756317
; Patent No. 5849894
; GENERAL INFORMATION:
; APPLICANT: Clemente, Thomas E.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Mitsky, Timothy A.
; APPLICANT: Stark, David M.
; TITLE OF INVENTION: Improved Rhodospirillum Rubrum
; TITLE OF INVENTION: Poly-B-Hydroxyalkonoate Synthase
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/756,317
; FILING DATE: 25-NOV-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,693
; FILING DATE: 29-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Patterson, Melinda L.
; REGISTRATION NUMBER: 33,062
; REFERENCE/DOCKET NUMBER: MOBT:008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 787-1400
; TELEFAX: (713) 787-1440
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 560 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-756-317-11

Query Match 2.4%; Score 110.5; DB 2; Length 560;
Best Local Similarity 19.7%; Pred. No. 0.032;
Matches 115; Conservative 82; Mismatches 216; Indels 171; Gaps 27;



```
                | |  |  :  ::|  :: :| |: |             |||  :
Db        445 FTV--AGSNDHITP--WDAVYRSALLLGGDRRFVLANSGHIQSI 484


RESULT 3
US-08-252-966B-12
; Sequence 12, Application US/08252966B
; Patent No. 5624818
;  GENERAL INFORMATION:
;    APPLICANT:  Eisenman, Robert N.
;    APPLICANT:  Hurlin, Peter J.
;    APPLICANT:  Ayer, Donald E.
;    TITLE OF INVENTION:  Regulatory Proteins that Dimerize with
;    TITLE OF INVENTION:  Mad or Max
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson, and KindnessPLLC
;      STREET:  1420 Fifth Ave., Suite 2800
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/252,966B
;      FILING DATE:  01-JUN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Shelton, Dennis K.
;      REGISTRATION NUMBER:  26,997.
;      REFERENCE/DOCKET NUMBER:  FHCR17694
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 682-8100
;      TELEFAX:  (206) 224-0779
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1253 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;      DESCRIPTION:  translation of msina cDNA; see Figure 23
;    HYPOTHETICAL:  YES
;    ORIGINAL SOURCE:
;      ORGANISM:  Mus musculus
US-08-252-966B-12

  Query Match             2.4%;  Score 108;  DB 1;  Length 1253;
  Best Local Similarity   18.5%;  Pred. No. 0.25;
  Matches  115;  Conservative  102;  Mismatches  215;  Indels  190;  Gaps   29;

Qy         98 LDHVKESKTGSSGIIVKLSTNHFRLTSR---------------PQWALYQYHIDYNPLME 142
              :|:    : |||   :  |    : | |               | |:          |
Db        565 IDYASCKRLGSSYRALPKSYQQPKCTGRTPLCKEVLNDTWVSFPSWS-----------E 612

Qy        143 ARRLRSALLFQHEDLIGRC--HAFDGTILFLPKRLQHKVTEVFSQ--TRNGEHVRITITL 198
                  |:   |:|: | ||    |:  ::        :| |   :  :|     :    |
Db        613 DSTFVSSKKTQYEEHIYRCEDERFELDVVLETNLATIRVLEAIQKKLSRLSAEEQAKFRL 672

Qy        199 TNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNY-----------YNPSDPIDIPNHR 246
              | |  ||              :::: : : | |           |||  : |   |
Db        673 DNTLGGTS--------------EVIHRKALQRIYADKAADIIDGLRKNPSIAVPIVLKR 717

Qy        247 LVI----W----PGFTTSILQYENNIMLCTDVSH----------KVLRSETVLDFMFNLY 288
              | :    |     ||   : : :|       : |          ||||:::|: : ::|
Db        718 LKMKEEEWREAQRGF-NKVWREQNEKYYLKSLDHQGINFKQNDTKVLRSKSLLNEIESIY 776

Qy        289 QQTEEHKFQE----------------QVSKELIGLIV--------LTKYNNKTYRVDDI 323
                : :|   :|                |: ::   ||:        : | :   |::  |
Db        777 DERQEQATEENAGVPVGPHLSLAYEDKQILEDAAALIIHHVKRQTGIQKEDK--YKIKQI 834

Qy        324 DWDQNPKSTF-KKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAM 382
                    |   | :: | |:|       ::  :|  |: :      ||:  | ||: |   :
Db        835 MHHFIPDLLFAQRGDLSDV-------EEEEEEEMDVDE--ATGAPKKHNGVGGSPPKSKL 885

Qy        383 XIPELC--YLTGLTDKMRNDFNVMKDLAVHTRLTP----------EQRQREVGRLIDYIH 430
                      | |: |:: | | |  :  :  ||             | :|::
Db        886 LFSNTAAQKLRGM-DEVYNLFYVNNNWYIFMRLHQILCLRLLRICSQAERQI-------- 936

Qy        431 KDDNVQRELRDW-----GLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWSKETRG 485
              :::| :||   |     |:  |    | |  |     |        || | | |  :
Db        937 EEENRERE---WEREVLGIKRDK---SDSPAIQLRLKEPMDVDVEDYYPAFLDMVRSLLD 990

Qy        486 APLISVKPLDNWLLIYTRRNYEA--ANSLIQNLFK-----VTPAMGIQMKKAIMIE---- 534
                : | :  |:   ::|   | |   : |||:: :     |:  : :|:    : |
Db        991 GNIDSSQYEDSLREMFTIHAYIAFTMDKLIQSIVRQLQHIVSDEVCVQVTDLYLAENNNG 1050

Qy        535 -----VDDRT------EAYLRALQQKVTSD------------TQIVVCLLSSNRKDK-- 568
                  :: :|       || |  :| :: :              |: | || :  ::
Db       1051 ATGGQLNSQTSRSLLESAYQRKAEQLMSDENCFKLMFIQSQGQVQLTVELLDTEEENSDD 1110

Qy        569 -------YDAIKKYLCTDCPTP 583
                    | :::|: :|  :|
Db       1111 PVEAERWSDYVERYMSSDTTSP 1132


RESULT 4
US-08-252-966B-18
; Sequence 18, Application US/08252966B
; Patent No. 5624818
;  GENERAL INFORMATION:
;    APPLICANT:  Eisenman, Robert N.
;    APPLICANT:  Hurlin, Peter J.
;    APPLICANT:  Ayer, Donald E.
;    TITLE OF INVENTION:  Regulatory Proteins that Dimerize with
;    TITLE OF INVENTION:  Mad or Max
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson, and KindnessPLLC
;      STREET:  1420 Fifth Ave., Suite 2800
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/252,966B
;      FILING DATE:  01-JUN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Shelton, Dennis K.
;      REGISTRATION NUMBER:  26,997.
;      REFERENCE/DOCKET NUMBER:  FHCR17694
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 682-8100
;      TELEFAX:  (206) 224-0779
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1261 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;      DESCRIPTION:  translation of msina9 cDNA; see Figure 29A, B, C, D
;    HYPOTHETICAL:  YES
;    ORIGINAL SOURCE:
;      ORGANISM:  Mus musculus
US-08-252-966B-18
```

```
  Query Match             2.4%;  Score 108;  DB 1;  Length 1261;
  Best Local Similarity   18.5%;  Pred. No. 0.25;
  Matches  115;  Conservative  102;  Mismatches  215;  Indels  190;  Gaps   29;

Qy         98 LDHVKESKTGSSGIIVKLSTNHFRLTSR---------------PQWALYQYHIDYNPLME 142
              :|:    : |||   :  |    : | |                | |:          |
Db        565 IDYASCKRLGSSYRALPKSYQQPKCTGRTPLCKEVLNDTWVSFPSWS-----------E 612

Qy        143 ARRLRSALLFQHEDLIGRC--HAFDGTILFLPKRLQHKVTEVFSQ--TRNGEHVRITITL 198
                  |:   |:|: | ||    |:  ::        :| |   :  :|     :    |
Db        613 DSTFVSSKKTQYEEHIYRCEDERFELDVVLETNLATIRVLEAIQKKLSRLSAEEQAKFRL 672

Qy        199 TNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNY-----------YNPSDPIDIPNHR 246
              | |  ||               :::: : : | |             |||  : |   |
Db        673 DNTLGGTS--------------EVIHRKALQRIYADKAADIIDGLRKNPSIAVPIVLKR 717

Qy        247 LVI----W----PGFTTSILQYENNIMLCTDVSH----------KVLRSETVLDFMFNLY 288
              | :    |     ||   : : :|       : |          |||||:::|: : ::|
Db        718 LKMKEEEWREAQRGF-NKVWREQNEKYYLKSLDHQGINFKQNDTKVLRSKSLLNEIESIY 776

Qy        289 QQTEEHKFQE----------------QVSKELIGLIV--------LTKYNNKTYRVDDI 323
              : :|   :|                |: ::   ||:        : | :   |::  |
Db        777 DERQEQATEENAGVPVGPHLSLAYEDKQILEDAAALIIHHVKRQTGIQKEDK--YKIKQI 834

Qy        324 DWDQNPKSTF-KKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAM 382
              |   | :: | |:|       ::  :|  |: :      ||:  | ||: |   :
Db        835 MHHFIPDLLFAQRGDLSDV-------EEEEEEEMDVDE--ATGAPKKHNGVGGSPPKSKL 885

Qy        383 XIPELC--YLTGLTDKMRNDFNVMKDLAVHTRLTP----------EQRQREVGRLIDYIH 430
                  | |: |:: | | |  :  :  ||             | :|::
Db        886 LFSNTAAQKLRGM-DEVYNLFYVNNNWYIFMRLHQILCLRLLRICSQAERQI-------- 936

Qy        431 KDDNVQRELRDW-----GLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWSKETRG 485
              :::| :||   |     |:  |    | |   |     |         || | | |  :
Db        937 EEENRERE---WEREVLGIKRDK---SDSPAIQLRLKEPMDVDVEDYYPAFLDMVRSLLD 990

Qy        486 APLISVKPLDNWLLIYTRRNYEA--ANSLIQNLFK-----VTPAMGIQMKKAIMIE---- 534
              : | :  |:   ::|   | |   : |||:: :     |:  : :|:    : |
Db        991 GNIDSSQYEDSLREMFTIHAYIAFTMDKLIQSIVRQLQHIVSDEVCVQVTDLYLAENNNG 1050

Qy        535 -----VDDRT------EAYLRALQQKVTSD-------------TQIVVCLLSSNRKDK-- 568
                  :: :|       || |  :| :: :               |: | || :  ::
Db       1051 ATGGQLNSQTSRSLLESAYQRKAEQLMSDENCFKLMFIQSQGQVQLTVELLDTEEENSDD 1110

Qy        569 -------YDAIKKYLCTDCPTP 583
                    | :::|: :|  :|
Db       1111 PVEAERWSDYVERYMSSDTTSP 1132


RESULT 5
US-08-844-280-2
; Sequence 2, Application US/08844280
; Patent No. 5747315
;  GENERAL INFORMATION:
;    APPLICANT:  Lawlor, Elizabeth
;    TITLE OF INVENTION:  No. 5747315el Compounds
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/844,280
;      FILING DATE:  18-APR-1997
;      CLASSIFICATION:  ?
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  9607993.4
;      FILING DATE:  18-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gimmi, Edward R
;      REGISTRATION NUMBER:  38,891
;      REFERENCE/DOCKET NUMBER:  P31457-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-4478
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  587 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-844-280-2

  Query Match             2.3%;  Score 106;  DB 1;  Length 587;
  Best Local Similarity   20.9%;  Pred. No. 0.1;
  Matches   72;  Conservative   59;  Mismatches  137;  Indels   76;  Gaps   15;

Qy        299 QVSKELIGLIVLTKYNN--KTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEIT 356
              |::|:|:      :|    | :| :|:  |: |:  | : |  | |||        |||
Db        198 QITKDLLMNAGFDRYYQIVKCFRDEDLRGDRQPE--FTQVD-LETSFLT------EQEIQ 248

Qy        357 DLKQPVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFN-VMKDLAVHTRLTP 415
              |: : ::    |  :|   ||| | |   :   | | :||    |: :::||
Db        249 DITEGLIARVMKETKGIEVTLPFPHMKYDDAMALYG-SDKPDTRFDMLLQDLT------- 300

Qy        416 EQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQ 475
                  || : :|:              |  |:    :::    :  |  |   :
Db        301 -----EVVKGVDF---------------KVFSEALAVKAIVVKGAADNYSRKDID---K 336

Qy        476 FADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEV 535
                : :|:      |  || :|  |       |   |  |      ::| |: :: |  ::
Db        337 MTEVAKQYGAKGLAWVKVVDGEL------NGPVAKFLTGIQEELTTALALEDKDLVLFVA 390

Qy        536 D--DRTEAYLRALQQKVTSDTQIVVCLLSSNRKDKYDAI--------------KKYLCTD 579
              |  :   | | ||: ::  :  ::          ||:: :             :|:
Db        391 DTLEVANATLGALRGRIAKELGLI-------DNDKFNFLWVVDWPMFEWSEEKSRYMSAH 443

Qy        580 CP--TPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVD 621
                |   | :       |    | |||  | |    ::||   |::
Db        444 HPFTLPQEETAHELKGDLAKVRAIAYDIVLN-GYELGGGSLRIN 486


RESULT 6
US-09-006-726-2
; Sequence 2, Application US/09006726
; Patent No. 6210940
; GENERAL INFORMATION:
;  APPLICANT: Lawlor, Elizabeth J.
;  TITLE OF INVENTION: No. 6210940el Compounds
;  FILE REFERENCE: P31457-2-D1
;  CURRENT APPLICATION NUMBER: US/09/006,726
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 08/844,280
;  EARLIER FILING DATE: 1997-04-18
;  EARLIER APPLICATION NUMBER: UK9607993.4
;  EARLIER FILING DATE: 1996-04-18
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 587
;   TYPE: PRT
;   ORGANISM: Streptococcus pneumoniae
US-09-006-726-2

  Query Match             2.3%;  Score 106;  DB 4;  Length 587;
```

Best Local Similarity 20.9%; Pred. No. 0.1;
Matches 72; Conservative 59; Mismatches 137; Indels 76; Gaps 15;

```
Qy        299 QVSKELIGLIVLTKYNN--KTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEIT 356
              |::|:|:      :|    | :| :|:  |: |:  | : |  | |||        |||
Db        198 QITKDLLMNAGFDRYYQIVKCFRDEDLRGDRQPE--FTQVD-LETSFLT------EQEIQ 248

Qy        357 DLKQPVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFN-VMKDLAVHTRLTP 415
              |: : ::   |  :|   ||| | |   :   | | :||    |: :::||
Db        249 DITEGLIARVMKETKGIEVTLPFPHMKYDDAMALYG-SDKPDTRFDMLLQDLT------- 300

Qy        416 EQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQ 475
                  || : :|:                 |  |:    :::    :  | |  :
Db        301 -----EVVKGVDF---------------KVFSEALAVKAIVVKGAADNYSRKDID---K 336

Qy        476 FADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEV 535
                : :|:     |  || :|  |      |   |  |     ::| |: :: |  ::
Db        337 MTEVAKQYGAKGLAWVKVVDGEL------NGPVAKFLTGIQEELTTALALEDKDLVLFVA 390

Qy        536 D--DRTEAYLRALQQKVTSDTQIVVCLLSSNRKDKYDAI-------------KKYLCTD 579
              |  :   | | ||: :: :  ::         ||:: :              :|:
Db        391 DTLEVANATLGALRGRIAKELGLI-------DNDKFNFLWVVDWPMFEWSEEKSRYMSAH 443

Qy        580 CP--TPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVD 621
                |  | :      |   | |||  | |    ::||   |::
Db        444 HPFTLPQEETAHELKGDLAKVRAIAYDIVLN-GYELGGGSLRIN 486
```

RESULT 7
US-09-224-024-28
; Sequence 28, Application US/09224024
; Patent No. 6056953
; GENERAL INFORMATION:
; APPLICANT: Leslie Hickle
; APPLICANT: Jewel Payne
; TITLE OF INVENTION: Materials and Methods for the Control of
; TITLE OF INVENTION: Calliphoridae Pests
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/224,024
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/856,226
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA79
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1180 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-224-024-28

Query Match 2.3%; Score 102.5; DB 3; Length 1180;
Best Local Similarity 17.2%; Pred. No. 0.82;
Matches 135; Conservative 112; Mismatches 242; Indels 295; Gaps 34;

```
Qy        187 RNGEHVRITITLT-----NELP------PTSPTCLQFYNI---------------------- 215
              :| : ||  | |       | :|      | :|:   :|||
Db        165 QNTQDVRTQIQLVHYHFQNVIPELVNSCPPNPSDCDYYNILVLSSYAQAANLHLTVLNQA 224

Qy        216 -XFRRLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSILQYEN------------ 262
               |   ||  | :|   :|  |      :|    :  :|  | :|  | |
Db        225 VKFEAYLK--NNRQF--DYLEP-----LPT-AIDYYPVLTKAIEDYTNYCVTTYKKGLNL 274

Qy        263 -----------NIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLT 311
                        ||   |  :::   :  |||  :  |:   :  |:   |  ||
Db        275 IKTTPDSNLDGNINWNTYNTYRTKMTTAVLD-LVALFPNYDVGKYPIGVQSEL------- 326

Qy        312 KYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRR 371
                : |:|  ::::::|               :||  || || :  ::| | :
Db        327 --TREIYQV--LNFEESP-------------YKYYDFQY-QEDSLTRRPHLFT------ 361

Qy        372 GPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHK 431
                          : | :          | |    ::  :|     |:    |
Db        362 -----------WLDSLNFYEKAQTTPNNFFTSHYNMFHYTLDNISQKSSVFG-------- 402

Qy        432 DDNVQRELRDWGLS-----FDSNLLSFSG--------------------RILQSEKIHQG 466
              : ||  :|:  ||:      |  |::|                        |:|: |
Db        403 NHNVTDKLKSLGLATNIYIFLLNVISLDNKYLNDYNNISKMDFFITNGTRLLEKELTAGS 462

Qy        467 GK-TFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGI 525
              |: |:| |       |   | |::  || :|  :      |:  : ::
Db        463 GQITYDVN-------KNIFGLPIL--KPRENQAIPTLFPTYDNYSHIL------------ 501

Qy        526 QMKKAIMIEVDDRTEAYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQ 585
                          ::::::|    |  ||:     : :  |  : |  :| | |
Db        502 --------------SFIKSLSIPATYKTQVYTFAWTHSSVDPKNTIYTHLTTQIP---- 542

Qy        586 CVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDT------- 638
              | | :||       |:|: :|     ||:|       |      | | |
Db        543 AVKANSLG-------TASKV-VQGPGHTGGDLIDFKDHFK------ITCQHSNFQQSYFI 588

Qy        639 --------TAGRRSIAGF----VASI---------------------------------- 652
                        :|  |::      || :
Db        589 RIRYASNGSANTRAVINLSIPGVAELGMALNPTFSGTDYTNLKYKDFQYLEFSNEVKFAP 648

Qy        653 NEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMP---------------- 695
              |: ::  |:|          |:| :              |::|
Db        649 NQNISLVFNRSDVYTNTTVLIDKI--------------EFLPITRSIREDREKQKLETV 693

Qy        696 SRVI--VYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQS 753
                ::|   | : : :      | :|:: |  : :: :      |:  :::: :  ||:  :||
Db        694 QQIINTFYANPIKNTLQSELTDYDIDQAANLVECISEELYPKEKMLLLDEVKNAKQLSQS 753

Qy        754 GGRLQN------PLPGTVID---VEVTRPEWYDFFIVSQAVR--SGSVSPTHYNVIYDSS 802
                |||       |  |  |   ::   | :   ::      |   |:: ||:     | |
Db        754 RNVLQNGDFESATLGWTTSDNITIQEDDPIFKGHYLHMSGAREIDGTIFPTYIFQKIDES 813

Qy        803 GLKP 806
                |||
Db        814 KLKP 817
```

RESULT 8
PCT-US94-07902-28
; Sequence 28, Application PC/TUS9407902
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT: Street address: 4980 Carroll Canyon Road
; APPLICANT: City: San Diego
; APPLICANT: State/Province: California
; APPLICANT: Country: US
; APPLICANT: Postal code/Zip: 92121
; APPLICANT: Phone number: (619) 453-8030 Fax number: . (619)453-6991
; APPLICANT: Telex number:

```
;    TITLE OF INVENTION:  Materials and Methods for the Control of
;    TITLE OF INVENTION:  Calliphoridae Pests
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David R. Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/07902
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, David R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  MA79
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  904-375-8100
;      TELEFAX:  904-372-5800
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1180 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US94-07902-28

  Query Match             2.3%;  Score 102.5;  DB 5;  Length 1180;
  Best Local Similarity   17.2%;  Pred. No. 0.82;
  Matches  135;  Conservative  112;  Mismatches  242;  Indels  295;  Gaps   34;

Qy        187 RNGEHVRITITLT-----NELP------PTSPTCLQFYNI---------- 215
              :| : ||  | |      | :|      | :|:   :|||
Db        165 QNTQDVRTQIQLVHYHFQNVIPELVNSCPPNPSDCDYYNILVLSSYAQAANLHLTVLNQA 224

Qy        216 -XFRRLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSILQYEN---------- 262
               |  || | :|   :| |      :|   :  :|  | :|  | |
Db        225 VKFEAYLK--NNRQF--DYLEP-----LPT-AIDYYPVLTKAIEDYTNYCVTTYKKGLNL 274

Qy        263 ----------NIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLT 311
                       ||   |  :::   :  ||| :  |:   :  |:   |  ||
Db        275 IKTTPDSNLDGNINWNTYNTYRTKMTTAVLD-LVALFPNYDVGKYPIGVQSEL------- 326

Qy        312 KYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRR 371
                 : |:|  :::::::|               :||  || || :  ::| | :
Db        327 --TREIYQV--LNFEESP-------------YKYYDFQY-QEDSLTRRPHLFT------ 361

Qy        372 GPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHK 431
                     : | :          | |    ::  :|     |:    |
Db        362 ----------WLDSLNFYEKAQTTPNNFFTSHYNMFHYTLDNISQKSSVFG-------- 402

Qy        432 DDNVQRELRDWGLS-----FDSNLLSFSG--------------------RILQSEKIHQG 466
              : ||  :|:  ||:     |  |::|                     |:|: |
Db        403 NHNVTDKLKSLGLATNIYIFLLNVISLDNKYLNDYNNISKMDFFITNGTRLLEKELTAGS 462

Qy        467 GK-TFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGI 525
              |: |:| |       |   | |::  || :|  :      |:  : ::
Db        463 GQITYDVN-------KNIFGLPIL--KPRENQAIPTLFPTYDNYSHIL------------ 501

Qy        526 QMKKAIMIEVDDRTEAYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQ 585
                      :::::|    |  ||:     :  |  :  | :|  | |  |
Db        502 --------------SFIKSLSIPATYKTQVYTFAWTHSSVDPKNTIYTHLTTQIP---- 542

Qy        586 CVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDT------- 638
              | | :||       |:|: :|     ||:|       |      | | |
```

```
Db        543 AVKANSLG-------TASKV-VQGPGHTGGDLIDFKDHFK------ITCQHSNFQQSYFI 588

Qy        639 --------TAGRRSIAGF----VASI---------------------------------- 652
                      :|  |::      || :
Db        589 RIRYASNGSANTRAVINLSIPGVAELGMALNPTFSGTDYTNLKYKDFQYLEFSNEVKFAP 648

Qy        653 NEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMP----------------- 695
              |: ::  |:|        |:| :               |::|
Db        649 NQNISLVFNRSDVYTNTTVLIDKI--------------EFLPITRSIREDREKQKLETV 693

Qy        696 SRVI--VYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQS 753
              ::|   | : : :     | :|:: |  : :: :     |:  :::: :  ||:  :||
Db        694 QQIINTFYANPIKNTLQSELTDYDIDQAANLVECISEELYPKEKMLLLDEVKNAKQLSQS 753

Qy        754 GGRLQN------PLPGTVID---VEVTRPEWYDFFIVSQAVR--SGSVSPTHYNVIYDSS 802
              |||       |  |  |   ::   | :   ::      |   |:: ||:     | |
Db        754 RNVLQNGDFESATLGWTTSDNITIQEDDPIFKGHYLHMSGAREIDGTIFPTYIFQKIDES 813

Qy        803 GLKP 806
              |||
Db        814 KLKP 817
```

```
RESULT 9
US-09-134-001C-3782
; Sequence 3782, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3782
;   LENGTH: 905
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3782

  Query Match             2.2%;  Score 101.5;  DB 4;  Length 905;
  Best Local Similarity   18.6%;  Pred. No. 0.65;
  Matches  126;  Conservative  102;  Mismatches  220;  Indels  229;  Gaps   36;

Qy        173 KRLQHK--VTEVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLK-IMNLQQI 229
              |:|: |  : :: : : | | :  |  :  :           | |:    | : ::: :
Db        304 KKLEFKQLLADIDTSSTNEEVIDKTFEIEQD----------FQNVDLNDLNEAVIHFELE 353

Qy        230 GRNYYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQ 289
              | ||   :          ::  ||      | |:         | |: :| | |:
Db        354 GTNYLKDT----------ILKFGF------YTNH-------QHVVINAEDVKDY------ 384

Qy        290 QTEEHKFQEQVSKELIGLIVLTKYNNKTY--------RVDDIDWDQNPKSTFKKADGSEV 341
              :|  |    |    ::   |   |||          :::|::|
Db        385 ---KHLVQWLEDKNTTKIVYDAK---KTYVSAHRLGINIENIEFD-------------- 423

Qy        342 SFLEYYRKQYNQEITDLKQPV-LVSQPKRR-------RGPGGTLPGPAMXIPELCYLT-- 391
              |  |    :: | |:|  | |  |   :        :|    :|   :    :  :|
Db        424 VMLASYIIDPSRSIDDVKSVVSLYGQNYVKDNITIFGKGKKHHIPEEPILNEHIASVTEA 483

Qy        392 --GLTDKMR------NDFNVMKDLAVH-TRLTPEQRQREVGRLIDYIHKDDNVQRELRDW 442
                :|  |:       |   ::|||.:   |:  |    |:|     |: | |   :|::
Db        484 IAAVTPTMKSQLEDYNQIELLKDLELPLARILSE--MEEIG-----IYTDIN---DLKEM 533

Qy        443 GLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYT 502
                  |     :| |     |: |: |      | :  | |     |   ::  |
Db        534 EFEIQKKL-----DVLISNIHESAGEAFNIN------SPKQLGVVLFETLQLP--VIKKT 580
```

```
Qy        503 RRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTE---------AYLRALQQKVTSD 553
Db        581 KTGYSTAVDVLEKLQGEHPI------------IDDILEYRQLAKLQSTYVEGLQKVISKD 628

Qy        554 ----TQIVVCLLSSNRKDKYD--------------AIKKYLCTDCPTPSQCVVARTLGKQ 595
Db        629 HRIHTRFNQTLAQTGRLSSIDPNLQNIPIRLEEGRKIRKAF---KPTSKDSVI------- 678

Qy        596 QTVMAIATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYH--------------DTTAG 641
Db        679 --LSADYSQIELRVLAHI-----TQDESLKHAFINGHDIHTATAMKVFNVESDQVDSLMR 731

Qy        642 RRSIA---GFVASINE-------GMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCN 691
Db        732 RQAKAVNFGIVYGISDYGLSQSLGITRKQAKAFIDD---------------YLASFPGVK 776

Qy        692 EYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFA 751
Db        777 QYMSD---IVKDAKAQGYVETLLH------------RRRYIPDIT----SRNVNLRSFA 816

Qy        752 QSGGRLQNPLPGTVIDV 768
Db        817 ERTA-MNTPIQGSAADI 832
```

RESULT 10
US-08-591-685-7
; Sequence 7, Application US/08591685
; Patent No. 6083733
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Thermostable xylanases
; NUMBER OF SEQUENCES: 13
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/591,685
; FILING DATE:
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-591-685-7

Query Match 2.2%; Score 99; DB 3; Length 351;
Best Local Similarity 20.1%; Pred. No. 0.22;
Matches 68; Conservative 52; Mismatches 123; Indels 96; Gaps 14;

```
Qy        222 KIMNLQQIGRNYY-----------NPSDPIDIPNHRLVIWP-------------GFTT 255
Db         19 KLPSLCEKYKNYFKIGVAVPYRALTNPVDVEMIKRHFNSITPENEMKPESLQPYEGGFSF 78

Qy        256 SIL-QY-----ENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQVSKELIGLI- 308
Db         79 SIADEYIDFCKKNNISL---RGHTLVWHQQTPSWFFTNPETGEKLTNSEKDKXILLDRLK 135

Qy        309 -----VLTKYNNKTYRVDDID--WDQNPKSTFKKADGSEVSFLEYYRKQ----------- 350
Db        136 KHIQTVVGRYKGKVYAWDVVNEAIDENQPDGYRRSDWYNIXGPEYIEKAFIWAHEADPKA 195

Qy        351 ---YNQEITD--LKQPVLVSQPKRRRGPGGTLPGPAMXIPELCYLT-------------- 391
Db        196 KLFYNDYSTEXPYKRGNLYTLIKNLKAKGVPVHGVGLQ----CHISLDWPDVSEIEETVK 251

Qy        392 ------GLTDKMRN-DFNVMKDLA---VHTRLTPEQRQREVGRLIDYIHKDDNVQRELRD 441
Db        252 LFSRIPGLEIHFTEIDISIAKNMTDDDAYNRYLLIQQAQKLKAIFDVLKKYRNVVTSVTF 311
```

```
Qy        442 WGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWS 480
Db        312 WGLKDDYSWLRGDMPLLFDK---------DYQPKFAFWS 341
```

RESULT 11
US-08-731-716-2
; Sequence 2, Application US/08731716
; Patent No. 5789202
; GENERAL INFORMATION:
; APPLICANT: Hoskins, JoAnn
; APPLICANT: Jaskunas, S. Richard
; APPLICANT: Rockey, Pamela K.
; APPLICANT: Zhao, Genshi
; APPLICANT: Rosteck, Paul R. Jr.
; APPLICANT: No. 5789202ris, Franklin H.
; TITLE OF INVENTION: Penicillin Binding Protein From
; TITLE OF INVENTION: Streptococcus Pneumoniae
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/731,716
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-10,887
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 731 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-731-716-2

Query Match 2.2%; Score 98.5; DB 1; Length 731;
Best Local Similarity 17.3%; Pred. No. 0.91;
Matches 114; Conservative 96; Mismatches 230; Indels 219; Gaps 25;

```
Qy         75 LSLAERGGRRRDFHDXGVNTRQNLDHVKESKTG-------------------SSGIIVK 114
Db         11 LSLFKKETSELEDSDSTILRRSRSDRKKLAQVGPIRKFWRRYHLTKIILILGLSAGLLV- 69

Qy        115 LSTNHFRLTSRPQWALYQYHI----DYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILF 170
Db         70 -------------GIYLFAVAKSTNVNDLQNALKTRTLIFDREEKEAGALSGQKGTYV- 114

Qy        171 LPKRLQHKVTEVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIG 230
Db        115 -------ELTDISKNLQN------AVIATED---------------------------- 132

Qy        231 RNYYNPSDPIDIPNHRLVIWP-----GFTTSILQYENNIMLCTDVSHKVLRSETVLDFMF 285
Db        133 RSFYK-NDGINYGRFFLAIVTAGRSGGGSTITQQLAKNAYLSQD--------------- 175

Qy        286 NLYQQTEEHKFQE-----QVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSE 340
Db        176 ----QTVERKAKEFFLALELSKKYSKEQILTMYLNNAYFGNGVWGVEDASKKYFGVSASE 231
```

```
Qy        341 VSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRND 400
Db        232 VS--------------LDQAATLA---------GMLKGPELYNP----LNSVEDSTNRR 263

Qy        401 FNVMKDL-------------AVHTRLTPEQRQREVGRLIDYIHK---DDNVQRELRDWGL 444
Db        264 DTVLQNMVAAGYIDKNQETEAAEVDMTSQLHDKYEGKISDYRYPSYFDAVVNEAVSKYNL 323

Qy        445 SFDSNLLSFSGRI-LQSEKIHQGGKTFDYN-----PQFADWSKETRGAPLISVKP----- 493
Db        324 T-EEEIVNNGYRIYTELDQNYQANMQIVYENTSLFPRAEDGTFAQSGSVALEPKTGGVRG 382

Qy        494 -----LDNWLLIYTRRNY-----EAANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTEAYL 543
Db        383 VVGQVADNDKTGFRNFNYATQSKRSPGSTIKPLVVYTPA--VEAGWALNKQLDNHTMQY- 439

Qy        544 RALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIAT 603
Db        440 ------------------DSYKVDNYAGIK--------TSREVPMYQSLAESLNLPAVAT 473

Qy        604 ------KIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGM 656
Db        474 VNDLGVDKAFEAGEKFGLNMEKVDRVLGVALGSGVE-----TNPLQMAQAYAAFANEGL 527
```

RESULT 12
US-09-308-375-2
; Sequence 2, Application US/09308375
; Patent No. 6300117
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394-PCT
; CURRENT APPLICATION NUMBER: US/09/308,375
; CURRENT FILING DATE: 1999-05-14
; EARLIER APPLICATION NUMBER: EP9719636.4
; EARLIER FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillius subtilis
US-09-308-375-2

Query Match 2.2%; Score 98.5; DB 4; Length 2285;
Best Local Similarity 19.2%; Pred. No. 6.9;
Matches 160; Conservative 140; Mismatches 333; Indels 199; Gaps 40;

```
Qy         16 QETVQHVGAAASQQPGYIPPRPQQSPTEGDLVGRGRQRGMVVGATSKSQELQISAGFQEL 75
Db        806 QKVKESRSLTSDEEQEYLQVTQQLAQTFPALVKGYDSQGNAILKTNKELEKAIENTKEYL 865

Qy         76 SLAE---RGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHFRLTSRPQWALYQ 132
Db        866 ALKKQETRDSAKKTFEDASKEIKKSKDELKQYK----------QIADYNDKGRPKWDLIA 915

Qy        133 YHIDYNPLMEARRLRSALLFQHEDL-IGRCHAFDGTI--------LFLPKRLQHKVTEVF 183
Db        916 DDDDYK--VAADKAKQSMLKAQSDIESGNAKVKDSVLSIANAYSSIDISNTLKTSISDVV 973

Qy        184 SQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIM----------------NL 226
Db        974 NK----------LNLKDDLDPEE---LEKFSSSLGKLQEKMQKALDSGDEKAFDNAKKDL 1020

Qy        227 QQIGRNYYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFN 286
Db       1021 QSLLETYSKSDSSIDV----------FKMSFDKAQKNIK-DGDKSLSSVKSE-VGDLGET 1068

Qy        287 LYQQTEEHKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEY 346
Db       1069 LAEAGNE---AEDFGKKL-------KEALDANSVDDI----------KAAIKEMSDAMQ 1107

Qy        347 YRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKD 406
Db       1108 F-----DSVQDVLNGDIFNNTKDQVAP---------------LNDLLEKMAEG----KS 1142

Qy        407 LAVHTRLTPEQRQREVGRLIDY------IHKDDNV-QRELRDWGLSFDSNLLSFSGRILQ 459
Db       1143 ISANEANTLIQKDKELAQAISIENGVVKINRDEVIKQRKVK---LDAYNDMVTYSNKLMK 1199

Qy        460 SEKIHQGGKTFDYNPQFADWSKETR--------GAPL--ISVKPLDNWLLIYTRRNYEA- 508
Db       1200 TE-VNNAIKTLNADTLRIDSLKKLRKERKLDMSEAELSDLEVKSINN--VADAKKELKKL 1256

Qy        509 ----------ANSLIQNLFKVTPAMGI-----------QMKKAIMIEVDDRTEAYLRAL 546
Db       1257 EEKMLQPGGYSNSQIEAMQSVKSALESYISASEEATSTQEMNKQALVEAGTSLENWTDQ- 1315

Qy        547 QQKVTSDTQIVVCLLSSNRK--DKYDA-IKKY--LCTDCPTPSQCVVARTLGKQQTVMAI 601
Db       1316 QEKANEETKTSMYVVDKYKEALEKVNAEIDKYNKQVNDYPKYSQ--------KYRD--AI 1365

Qy        602 ATKI-ALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWF 660
Db       1366 KKEIKALQQKKKLMQEQAKL---LKDQIKSGNITQYGIVTSTTSSGGTPSSTGGSYSGKY 1422

Qy        661 SRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQ 720
Db       1423 SSYINSAASKYNVD--PALIAAVIQQESGFNAKA-------RSGVGAMGLMQL----MPA 1469

Qy        721 FLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPL------PGTVI 766
Db       1470 TAKSL-GVNNAYDPYQNVMGGTKYL-AQQLEKFGGNVEKALAAYNAGPGNVI 1519
```

RESULT 13
US-09-134-001C-3636
; Sequence 3636, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3636
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3636

Query Match 2.2%; Score 98; DB 4; Length 600;
Best Local Similarity 19.9%; Pred. No. 0.72;
Matches 113; Conservative 92; Mismatches 182; Indels 182; Gaps 34;

```
Qy         98 LDHVKESKTGSSGIIVKLSTNHFRLTSRPQ--WALYQYHIDYNP---LMEARRLRSALLF 152
Db         10 IDEIK-NKTD----ILDLVSEYVKLEKRGRNYIGLCPFHDEKTPSFTVSEDKQI------ 58

Qy        153 QHEDLIGRCHAF----DGTILFLPKRLQHKVTEVFSQTRNGEHVRITITLTNELPPTSPT 208
Db         59 --------CHCFGCKKGGNVFQFTQEIK-DVSFVEAVKDLGERVNIQVDIGQNQTNSSTK 109

Qy        209 CLQFYNIXFRRLLKIMNLQQIGRNYYN--------PSDPIDIPNHRLVIWPGFTTSILQY 260
Db        110 IAS-------DELKMIEMHELIKDYYHYALMKTVEGEEALNYLHER-----GFTDDLIK- 156

Qy        261 ENNIMLCTDVSH------KVLRSETVLDFMFNLYQQTEEH-----KFQEQVS---KELIG 306
```

```
Db        157 EREIGYAPDNSHFCHDFLEKKGYDIELAFEAGLLSRNEENFTYFDRFRNRIMFPLKNGQG 216

Qy        307 LIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDL--KQPVLV 364

Db        217 RIV--GYSGRTYT------DQEP-------------------KYLNSPETPIFQKRRILY 249

Qy        365 SQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDL--AVHTRLTPEQRQREV 422

Db        250 NLNKARKF--------IRKQDEIILLEGFMDVIKSDYAGLKQVVASMGTQLSQEH----- 296

Qy        423 GRLIDYIHK-DDNVQRELRDWGLSFDSNLLSFSGRILQSEK-IHQGGKTF---------- 470

Db        297 ---ITFLQKLTQNVT-------LMFDGDYAGKEATLKTGQALLNQGLNVYVVQLPSGMDP 346

Qy        471 -DY-----NPQFADW-SKETRGAPLISVKPLDNWL----LIYTRRNYEAA---------- 509

Db        347 DDYIRKYDNEQFLKFVQQDKQSFVLFKVKMYQNEINHNDLAYEKHFKETVRDLSLVNSGI 406

Qy        510 --NSLIQN---LFKVTPAMGIQMKKAIMIEVDDRTEAYLRALQQKVTSDT---------- 554

Db        407 IRNKLIQNIADIFKVNP-------ETIQYELD-------ATYQHQMSSNTYPTFQDEPSK 452

Qy        555 -QIVVCLLSSNRKDKYDAIKKYLCTDCPT 582

Db        453 QQLILGRLTKNEKAE-RALIKHLTKDKDT 480


RESULT 14
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
;   LENGTH: 10182
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

  Query Match             2.2%;  Score 98;  DB 4;  Length 10182;
  Best Local Similarity   19.0%;  Pred. No. 1.1e+02;
  Matches  118;  Conservative  101;  Mismatches  234;  Indels  168;  Gaps   27;

Qy         59 ATSKSQEL-----------QISAGFQELSLAERG----GRRRDFHDXGVNTRQNLDHVK 102

Db       6866 AIAKGQEIINSENNPTISSTDINRTIQEINDAEQNLHGDNKLRQAQEIAKNEIQNLDGL- 6924

Qy        103 ESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCH 162

Db       6925 -----NSAQITKLIQDIGRTTTKP--AVTQ------KLEEAKAINQAM----------- 6959

Qy        163 AFDGTILFLPKRLQHKVTEVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLK 222

Db       6960 ----------QQLKQSIADKDA----------TLNSSNYLNEDSEKKLAYDN-------A 6992

Qy        223 IMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLD 282

Db       6993 VSQAEQLINQLNDPT--MDISNIQAI-----TQKVIQAKDSLHGANKLAQNQADSNLIIN 7045

Qy        283 FMFNL------------YQQTEEHKFQ-----EQVSKELIGLIVL----------TKYN 314

Db       7046 QSTNLNDKQKQALNDLINHAQTKQQVAEIIAQANKLNNEMGTLKTLVEEQSNVHQQSKYI 7105

Qy        315 NKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPG 374

Db       7106 NEDPQVQNIYNDSIQKGR-EILNGTTDDVLN------NNKIADAIQNIHLTK-NDLHGDQ 7157

Qy        375 GTLPGPAMXIPELCYLTGLTDKMR----NDFN-------VMKDLAVHTRLTPEQRQRE-- 421

Db       7158 KLQKAQQDATNELNYLTNLNNSQRQSEHDEINSAPSRTEVSNDLNHAKALNEAMRQLENE 7217

Qy        422 ------VGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQG--GKTFDYN 473

Db       7218 VALENSVKKLSDFINEDEAAQNE-------------YSNALQKAKDIINGVPSSTLD-- 7261

Qy        474 PQFADWSKETRGAPLISV----KPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKK 529

Db       7262 -------KATIEDALLELQNARESLHGEQKLQEAKNQAVAE--IDNLQALNPGQ-VLAEK 7311

Qy        530 AIMIEVDDRTEAYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVA 589

Db       7312 TLVNQASTKPEVQ-EALQKAKELNEAMKALKTEINKKEQIKADSRYVNADSGLQANYNSA 7370

Qy        590 RTLGKQQTVMAIATKIALQMN 610

Db       7371 LNYGSQ----IIATTQPPELN 7387


RESULT 15
US-08-996-441B-6
; Sequence 6, Application US/08996441B
; Patent No. 6023013
;  GENERAL INFORMATION:
;    APPLICANT:  English, Leigh H.
;    APPLICANT:  Brussock, Susan M.
;    APPLICANT:  Malvar, Thomas M.
;    APPLICANT:  Bryson, James W.
;    APPLICANT:  Kulesza, Caroline A.
;    APPLICANT:  Walters, Frederick S.
;    APPLICANT:  Slatin, Stephen L.
;    APPLICANT:  Von Tersch, Michael A.
;    APPLICANT:  Romano, Charles
;    TITLE OF INVENTION:  INSECT-RESISTANT TRANSGENIC PLANTS
;    NUMBER OF SEQUENCES:  113
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/996,441B
;      FILING DATE:  18-DEC-1997
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kitchell, Barbara S.
;      REGISTRATION NUMBER:  33,928
;      REFERENCE/DOCKET NUMBER:  MECO:151
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  512/418-3000
;      TELEFAX:  512/474-7577
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  652 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-996-441B-6
```

Query Match 2.2%; Score 97.5; DB 3; Length 652;
Best Local Similarity 20.2%; Pred. No. 0.94;
Matches 69; Conservative 49; Mismatches 115; Indels 109; Gaps 16;

```
Qy        155 EDLIGRCHAFDGTILFL-PKRLQHKVTEVFSQTRNGEHVRITITLTNELPPTS------- 206
              || :   :::  | | |  || | :: |:|||  :  | |      | :|  :
Db        142 EDYVNALNSWKKTPLSLRSKRSQDRIRELFSQAES--HFR------NSMPSFAVSKFEVL 193

Qy        207 --PTCLQFYNIXFRRLLKIMNLQQIGRNY-YNPSDPIDIPNHRLVIWPGFTTSILQYENN 263
                ||  |  |     || : : |  |  : |:| |  :  : :| :  :|   : : |
Db        194 FLPTYAQAAN---THLLLLKDAQVFGEEWGYSPEDVAEFSHRQLKLTQQYTDHCVNWYN- 249

Qy        264 IMLCTDVSHKVLRSETVLDFM-FNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTYRVDD 322
                  |    ||  |   :: ||        :|: :::  :: ||||  : :
Db        250 ------VGLNGLRGSTYDAWVKFN--------RFRREMTLTVLDLIVLFPFYD------- 288

Qy        323 IDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITD--LKQPVLVSQPKRRRGPGGTLPGP 380
                           :  | |    |:|      |:      :  ||
Db        289 ---------------------IRLYSKGVKTELTRDIFTDPIFSLNTLQEYGPTFLSIEN 327

Qy        381 AMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELR 440
              ::  | |             |: :: :  |||| |           | ||
Db        328 SIRKPHL------------FDYLQGIEFHTRLQP----------GYFGKD-------- 355

Qy        441 DWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWSKE 482
                ||:     :||  :::       ||   :| : | | |
Db        356 ----SFN----YWSGNYVETRPSIGSSKTIT-SPFYGDKSTE 388
```

Search completed: July 22, 2003, 07:06:35
Job time : 13.4217 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 07:01:08 ; Search time 22.1715 Seconds
(without alignments)
4617.236 Million cell updates/sec

Title: US-09-873-737A-4
Perfect score: 4528
Sequence: 1 MTGRARARARGRARGQETVQ.....VGQSIHREPNLSLSNRLYYL 862

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep1:*
11: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep2:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep3:*
13: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4520	99.8	862	10	US-09-873-737A-4	Sequence 4, Appli
2	4286.5	94.7	861	10	US-09-873-737A-6	Sequence 6, Appli
3	3957	87.4	775	14	US-10-043-774B-2	Sequence 2, Appli
4	2687	59.3	523	11	US-09-998-598-2603	Sequence 2603, Ap
5	2687	59.3	523	15	US-10-029-495-3	Sequence 3, Appli
6	1459.5	32.2	866	15	US-10-029-495-4	Sequence 4, Appli
7	1406	31.1	844	14	US-10-043-774B-13	Sequence 13, Appl
8	1404	31.0	844	15	US-10-029-495-7	Sequence 7, Appli
9	1385.5	30.6	843	10	US-09-873-737A-2	Sequence 2, Appli
10	1360.5	30.0	580	10	US-09-801-574-16	Sequence 16, Appl
11	1299.5	28.7	498	15	US-10-029-495-9	Sequence 9, Appli
12	1290	28.5	824	15	US-10-029-495-6	Sequence 6, Appli
13	1137	25.1	722	15	US-10-029-495-5	Sequence 5, Appli
14	620	13.7	130	15	US-10-102-806-685	Sequence 685, App
15	511	11.3	1101	15	US-10-174-363-54	Sequence 54, Appl
16	499.5	11.0	915	15	US-10-174-363-40	Sequence 40, Appl
17	499.5	11.0	1048	15	US-10-174-363-56	Sequence 56, Appl
18	497	11.0	1085	15	US-10-174-363-12	Sequence 12, Appl
19	496	11.0	1073	15	US-10-174-363-14	Sequence 14, Appl
20	490	10.8	186	15	US-10-029-495-2	Sequence 2, Appli
21	490	10.8	1058	15	US-10-174-363-38	Sequence 38, Appl
22	486.5	10.7	1065	15	US-10-174-363-28	Sequence 28, Appl
23	457	10.1	901	15	US-10-174-363-8	Sequence 8, Appli
24	457	10.1	904	15	US-10-174-363-55	Sequence 55, Appl
25	455	10.0	969	12	US-09-533-029-106	Sequence 106, App
26	455	10.0	969	15	US-10-286-264-102	Sequence 102, App
27	454	10.0	821	15	US-10-174-363-4	Sequence 4, Appli
28	449	9.9	723	15	US-10-174-363-24	Sequence 24, Appl
29	444.5	9.8	757	15	US-10-174-363-10	Sequence 10, Appl
30	440.5	9.7	916	15	US-10-174-363-42	Sequence 42, Appl
31	429.5	9.5	1145	11	US-09-866-557A-5	Sequence 5, Appli
32	429.5	9.5	1145	15	US-10-055-797-5	Sequence 5, Appli
33	427.5	9.4	910	15	US-10-174-363-22	Sequence 22, Appl
34	399	8.8	577	15	US-10-174-363-2	Sequence 2, Appli
35	391	8.6	919	15	US-10-174-363-16	Sequence 16, Appl
36	368.5	8.1	442	15	US-10-174-363-26	Sequence 26, Appl
37	350	7.7	411	15	US-10-174-363-6	Sequence 6, Appli
38	339	7.5	389	15	US-10-174-363-50	Sequence 50, Appl
39	305.5	6.7	346	15	US-10-174-363-32	Sequence 32, Appl
40	210.5	4.6	71	10	US-09-864-761-48051	Sequence 48051, A
41	151.5	3.3	1924	11	US-09-866-557A-2	Sequence 2, Appli
42	151.5	3.3	1924	15	US-10-055-797-2	Sequence 2, Appli
43	144	3.2	55	10	US-09-864-761-38135	Sequence 38135, A
44	140	3.1	193	15	US-10-174-363-44	Sequence 44, Appl
45	139	3.1	322	15	US-10-174-363-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-09-873-737A-4
; Sequence 4, Application US/09873737A
; Patent No. US20020076797A1
; GENERAL INFORMATION:
; APPLICANT: Duke University
; APPLICANT: Lin, Haifan
; TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
; TITLE OF INVENTION: PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
; FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2
; CURRENT APPLICATION NUMBER: US/09/873,737A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/US99/28764
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/110,901
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Mus sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (90)
; OTHER INFORMATION: Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (216)
; OTHER INFORMATION: Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (383)
; OTHER INFORMATION: Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (816)
; OTHER INFORMATION: Xaa=Leu or Ile
US-09-873-737A-4

Query Match 99.8%; Score 4520; DB 10; Length 862;
Best Local Similarity 100.0%; Pred. No. 0;

```
  Matches  862;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEGDLVGRGRQRGMVVGAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEGDLVGRGRQRGMVVGAT 60

Qy         61 SKSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHF 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHF 120

Qy        121 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVT 180

Qy        181 EVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 EVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPI 240

Qy        241 DIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 DIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQV 300

Qy        301 SKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQ 360

Qy        361 PVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 PVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 420

Qy        421 EVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWS 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWS 480

Qy        481 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTE 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTE 540

Qy        541 AYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 600

Qy        601 IATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWF 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWF 660

Qy        661 SRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQ 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 SRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQ 720

Qy        721 FLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 FLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 780

Qy        781 VSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 VSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLA 840

Qy        841 FLVGQSIHREPNLSLSNRLYYL 862
              ||||||||||||||||||||||
Db        841 FLVGQSIHREPNLSLSNRLYYL 862


RESULT 2
US-09-873-737A-6
; Sequence 6, Application US/09873737A
; Patent No. US20020076797A1
; GENERAL INFORMATION:
;  APPLICANT: Duke University
;  APPLICANT:  Lin, Haifan
;  TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
;  TITLE OF INVENTION:  PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
;  FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2
;  CURRENT APPLICATION NUMBER: US/09/873,737A
;  CURRENT FILING DATE:  2001-06-04
;  PRIOR APPLICATION NUMBER: PCT/US99/28764
;  PRIOR FILING DATE: 1999-12-03
;  PRIOR APPLICATION NUMBER: 60/110,901
;  PRIOR FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;  LENGTH: 861
;  TYPE: PRT
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (76)
;  OTHER INFORMATION: Xaa=Leu or Ile
;  NAME/KEY: misc_feature
;  LOCATION: (303)
;  OTHER INFORMATION: Xaa=Leu or Ile
;  NAME/KEY: misc_feature
;  LOCATION: (735)
;  OTHER INFORMATION: Xaa=Leu or Ile
US-09-873-737A-6

  Query Match             94.7%;  Score 4286.5;  DB 10;  Length 861;
  Best Local Similarity   93.7%;  Pred. No. 0;
  Matches  808;  Conservative   22;  Mismatches   31;  Indels    1;  Gaps    1;

Qy          1 MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEGDLVGRGRQRGMVVGAT 60
              |||||||||||||||||| | ||: |||||||| |||| | ||:| ||||||| | |
Db          1 MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRG-TAGGT 59

Qy         61 SKSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHF 120
              :||| |||||||||| |||||||||| |||||||||||||||||||||:||||||
Db         60 AKSQGLQISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHF 119

Qy        121 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVT 180
              |||||||||||||||||||||||||||||||||||||:|||||||||||||||| |||
Db        120 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVT 179

Qy        181 EVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPI 240
              ||||:||||| |||||||||||||||||||| ||||||||||||||||||:|||
Db        180 EVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPI 239

Qy        241 DIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQV 300
              |||:||||||||||||||:|||||||||||||||||||| | ||||||||||
Db        240 DIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQV 299

Qy        301 SKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQ 360
              ||| |||:|||||| |||||||||||||||||||||| ||||||||||||||||||||
Db        300 SKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQ 359

Qy        361 PVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 420
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        360 PVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 419

Qy        421 EVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWS 480
              ||||||||||:|||||||||||||||||||||||||||:|||||||||||||||||||||
Db        420 EVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWS 479

Qy        481 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTE 540
              ||||||||||||||||||||||||||||||||||||||||||||:||:||||||||||||
Db        480 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDRTE 539

Qy        541 AYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 600
              |||| |||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 AYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 599

Qy        601 IATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWF 660
              |||||||||||||||||||: ||| |||||||||| ||||||||||||||||||||||||
Db        600 IATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWF 659

Qy        661 SRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQ 720
```

```
                ||:|||||||||||||||||||||||: |||||||:|||||||||||||||||||||||
Db        660 SRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQ 719

Qy        721 FLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 780
                |||||||:|||||||  ||||||||||  ||||||||||||||||||||||||||||||||
Db        720 FLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 779

Qy        781 VSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLA 840
                |||||||||||||||||||:|||||||||||||| ||:||||||||||||||||||||||
Db        780 VSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLA 839

Qy        841 FLVGQSIHREPNLSLSNRLYYL 862
                |||||||||||||||||||||
Db        840 FLVGQSIHREPNLSLSNRLYYL 861


RESULT 3
US-10-043-774B-2
; Sequence 2, Application US/10043774B
; Publication No. US20030129599A1
; GENERAL INFORMATION:
;  APPLICANT: University of Illinois at Chicago
;  APPLICANT:  Sharma, Arun
;  APPLICANT:  Hoffman, Ronald
;  TITLE OF INVENTION: HUMAN HEMATOPOIETIC GROWTH REGULATORY GENE AND USES
;  FILE REFERENCE: MBHB: CU08/PPA
;  CURRENT APPLICATION NUMBER: US/10/043,774B
;  CURRENT FILING DATE:  2002-06-06
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 775
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-043-774B-2

  Query Match            87.4%;  Score 3957;  DB 14;  Length 775;
  Best Local Similarity  95.9%;  Pred. No. 0;
  Matches  740;  Conservative   19;  Mismatches   13;  Indels    0;  Gaps    0;

Qy         91 GVNTRQNLDHVKESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSAL 150
                |||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||
Db          4 GVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSAL 63

Qy        151 LFQHEDLIGRCHAFDGTILFLPKRLQHKVTEVFSQTRNGEHVRITITLTNELPPTSPTCL 210
                |||||||||:|||||||||||||||| |||||||:|||| ||||||||||||||||||||
Db         64 LFQHEDLIGKCHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCL 123

Qy        211 QFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDV 270
                ||||| |||||||||||||||||:||||||:|||||||||||||||:|||||||||||||
Db        124 QFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDV 183

Qy        271 SHKVLRSETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPK 330
                |||||||||||||||| | ||||||||||||||||:||||||||||||||||||||||||
Db        184 SHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKELIGLVVLTKYNNKTYRVDDIDWDQNPK 243

Qy        331 STFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMXIPELCYL 390
                |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        244 STFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYL 303

Qy        391 TGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNL 450
                |||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||
Db        304 TGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNL 363

Qy        451 LSFSGRILQSEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAAN 510
                |||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 LSFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAAN 423

Qy        511 SLIQNLFKVTPAMGIQMKKAIMIEVDDRTEAYLRALQQKVTSDTQIVVCLLSSNRKDKYD 570
                |||||||||||||:||:|||||||||||||||| ||||||:|||||||||||||||||||
Db        424 SLIQNLFKVTPAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYD 483

Qy        571 AIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMALKLAMIV 630
                |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        484 AIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIV 543

Qy        631 GIDCYHDTTAGRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGC 690
                ||||||| |||||||||||||||||||||||||:|||||||||||||||||||||||||| |
Db        544 GIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSC 603

Qy        691 NEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFF 750
                |||||||:|||||||||||||||||||||||||||||:|||||||||||||||||| |||
Db        604 NEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVVKKRVNTRFF 663

Qy        751 AQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQ 810
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db        664 AQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQ 723

Qy        811 RLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 862
                ||||| ||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 RLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 775


RESULT 4
US-09-998-598-2603
; Sequence 2603, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
;  APPLICANT: Stolk, John A.
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Chenault, Ruth A.
;  APPLICANT:  Meagher, Madelein Joy
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.561
;  CURRENT APPLICATION NUMBER: US/09/998,598
;  CURRENT FILING DATE:  2001-11-16
;  NUMBER OF SEQ ID NOS: 2606
;  SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 2603
;   LENGTH: 523
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-998-598-2603

  Query Match            59.3%;  Score 2687;  DB 11;  Length 523;
  Best Local Similarity  96.2%;  Pred. No. 8.6e-256;
  Matches  502;  Conservative   12;  Mismatches    8;  Indels    0;  Gaps    0;

Qy        341 VSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRND 400
                |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db          2 VSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRND 61

Qy        401 FNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQS 460
                |||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||:
Db         62 FNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQT 121

Qy        461 EKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVT 520
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 EKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVT 181

Qy        521 PAMGIQMKKAIMIEVDDRTEAYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDC 580
                ||||:||:|||||||||||| |||||:|||||||||||||||||||||||||||||||||
Db        182 PAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDC 241

Qy        581 PTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTA 640
                |||||||||||||||||||||||||||||||||||||||||: ||| |||||||||| ||
Db        242 PTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTA 301

Qy        641 GRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIV 700
                |||||||||||||||||||:|||||||||||||||||||||||: ||||||||:||
Db        302 GRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIV 361

Qy        701 YRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNP 760
                |||||||||||||||||||||||:|||||||||||||||||| ||||||||||||||||
Db        362 YRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVVKKRVNTRFFAQSGGRLQNP 421
```

```
Qy        761 LPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVY 820
              ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||| ||:|
Db        422 LPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIY 481

Qy        821 YNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 862
              |||||||||||||||||||||||||||||||||||||||||
Db        482 YNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 523


RESULT 5
US-10-029-495-3
; Sequence 3, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
;  APPLICANT: Drmanac, Radoje T.
;  APPLICANT:  Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Dickson, Mark C.
;  APPLICANT:  Jones, Lee William
;  APPLICANT:  Ballinger, Dennis G.
;  APPLICANT:  Xue, Aidong
;  APPLICANT:  Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi,Vinod
;  TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
;  FILE REFERENCE: 28110/35905A
;  CURRENT APPLICATION NUMBER: US/10/029,495
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: 09/378,667
;  PRIOR FILING DATE: 1999-08-20
;  PRIOR APPLICATION NUMBER: 09/687,527
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 523
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: hiwi
US-10-029-495-3

  Query Match             59.3%;  Score 2687;  DB 15;  Length 523;
  Best Local Similarity   96.2%;  Pred. No. 8.6e-256;
  Matches  502;  Conservative   12;  Mismatches    8;  Indels    0;  Gaps    0;

Qy        341 VSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRND 400
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db          2 VSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRND 61

Qy        401 FNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQS 460
              ||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||:
Db         62 FNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQT 121

Qy        461 EKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVT 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 EKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVT 181

Qy        521 PAMGIQMKKAIMIEVDDRTEAYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDC 580
              ||||:||:||||||||||||||| |||||:||||||||||||||||||||||||||||||
Db        182 PAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDC 241

Qy        581 PTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTA 640
              |||||||||||||||||||||||||||||||||||||: ||| ||||||||| ||
Db        242 PTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTA 301

Qy        641 GRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIV 700
              |||||||||||||||||||:|||||||||||||||||||||: ||||||||:||
Db        302 GRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIV 361

Qy        701 YRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNP 760
              |||||||||||||||||||||||||:|||||||||||||||| ||||||||||||
Db        362 YRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVVKKRVNTRFFAQSGGRLQNP 421

Qy        761 LPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVY 820
              ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||| ||:|
Db        422 LPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIY 481

Qy        821 YNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 862
              |||||||||||||||||||||||||||||||||||||||||
Db        482 YNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 523


RESULT 6
US-10-029-495-4
; Sequence 4, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
;  APPLICANT: Drmanac, Radoje T.
;  APPLICANT:  Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Dickson, Mark C.
;  APPLICANT:  Jones, Lee William
;  APPLICANT:  Ballinger, Dennis G.
;  APPLICANT:  Xue, Aidong
;  APPLICANT:  Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi,Vinod
;  TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
;  FILE REFERENCE: 28110/35905A
;  CURRENT APPLICATION NUMBER: US/10/029,495
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: 09/378,667
;  PRIOR FILING DATE: 1999-08-20
;  PRIOR APPLICATION NUMBER: 09/687,527
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 866
;   TYPE: PRT
;   ORGANISM: Drosophila
;   FEATURE:
;   OTHER INFORMATION: aubergine
US-10-029-495-4

  Query Match             32.2%;  Score 1459.5;  DB 15;  Length 866;
  Best Local Similarity   36.4%;  Pred. No. 2.5e-134;
  Matches  320;  Conservative  174;  Mismatches  339;  Indels   47;  Gaps   18;

Qy          7 ARARGRARGQETVQHVGAAASQQPGYIPPRPQQS-PTEGDLVGRGRQRGMVVGATSKSQE 65
              || ||| |    |:       .  |: |   |:| |:  :    | |     |      :
Db         10 ARGRGRGRKPNNVE-------ANRGFAPSLGQKSDPSHSE----GNQAS---GGNGGGGD 55

Qy         66 LQISAGFQELSLAERGGRRRDFHDXG-VNTRQNLDHV-------KESKTGSSGIIVKLST 117
               |:    :: ||:     : :    | |  |: :  :          || |  |  : :
Db         56 AQVGPSIEKSSLSAVQMHKSEGDPRGSVRGRRLITDLVYSRPPGMTSKKGVVGTHITVQA 115

Qy        118 NHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLP---KR 174
              |:|::  || | :||| :|: | :|| ||| : |::|: ::|  : |||| :|     |
Db        116 NYFKVLKRPNWTIYQYRVDFTPDVEATRLRRSFLYEHKGILGG-YIFDGTNMFCINQFKA 174

Qy        175 LQHK--VTEVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRN 232
              :|    | |: :::| ||:: | |     :  |      |  |:  || :: ::|: : |
Db        175 VQDSPYVLELVTKSRAGENIEIKIKAVGSVQSTDAEQFQVLNLILRRAMEGLDLKLVSRY 234

Qy        233 YYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTE 292
              ||:|    |:: | |: :|||: ||| |:||:|:||::: |||:|:||: : : :  : ::
Db        235 YYDPQAKINLENFRMQLWPGYQTSIRQHENDILLCSEICHKVMRTETLYNILSDAIRDSD 294

Qy        293 EHKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYN 352
              :  :|    : ::|:::|| |||||||:||:|:   |   ||  || |:|:::||:|:||
Db        295 D--YQSTFKRAVMGMVILTDYNNKTYRIDDVDFQSTPLCKFKTNDG-EISYVDYYKKRYN 351

Qy        353 QEITDLKQPVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTR 412
```

```
Db        352 IIIRDLKQPLVMSRPTDKNIRGGN-DQAIMIIPELARATGMTDAMRADFRTLRAMSEHTR 410

Qy        413 LTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDY 472

Db        411 LNPDRRIERLRMFNKRLKSCKQSVETLKSWNIELDSALVEIPARVLPPEKILFGNQKIFV 470

Qy        473 NPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIM 532

Db        471 CDARADWTNEFRTCSMFKNVHINRWYVITPSRNLRETQEFVQMCIRTASSMKMNICNPIY 530

Qy        533 IEV-DDRTEAYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVART 591

Db        531 EEIPDDRNGTYSQAIDNAAANDPQIVMVVMRSPNEEKYSCIKKRTCVDRPVPSQVVTLKV 590

Qy        592 LGKQQT----VMAIATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAG 647

Db        591 IAPRQQKPTGLMSIATKVVIQMNAKLMGAPWQVVIPLHGLMTVGFDVCHSPKNKNKSYGA 650

Qy        648 FVASINEGMT-RWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVG 706

Db        651 FVATMDQKESFRYFSTVNEHIKGQELSEQMSVNMACALRSYQEQHRSLPERILFFRDGVG 710

Qy        707 DGQLKTLVNYEVPQFLDCL----KSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLP 762

Db        711 DGQLYQVVNSEVNTLKDRLDEIYKSAGKQEGCRMTFIIVSKRINSRYF--TGHR--NPVP 766

Qy        763 GTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYN 822

Db        767 GTVVDDVITLPERYDFFLVSQAVRIGTVSPTSYNVISDNMGLNADKLQMLSYKMTHMYYN 826

Qy        823 WPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 862

Db        827 YSGTIRVPAVCHYAHKLAFLVAESINRAPSAGLQNQLYFL 866
```

RESULT 7
US-10-043-774B-13
; Sequence 13, Application US/10043774B
; Publication No. US20030129599A1
; GENERAL INFORMATION:
; APPLICANT: University of Illinois at Chicago
; APPLICANT: Sharma, Arun
; APPLICANT: Hoffman, Ronald
; TITLE OF INVENTION: HUMAN HEMATOPOIETIC GROWTH REGULATORY GENE AND USES
; FILE REFERENCE: MBHB: CU08/PPA
; CURRENT APPLICATION NUMBER: US/10/043,774B
; CURRENT FILING DATE: 2002-06-06
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 844
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: PIWI protein
US-10-043-774B-13

Query Match 31.1%; Score 1406; DB 14; Length 844;
Best Local Similarity 36.0%; Pred. No. 4.4e-129;
Matches 308; Conservative 173; Mismatches 317; Indels 58; Gaps 18;

```
Qy         48 GRGRQRGMVVGATSKSQ--------------------ELQISAGFQELSLAE----RGG 82

Db          6 GRGRRRPLNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG 65

Qy         83 --RRRDFHD--XGVNTRQNLDHVKESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYN 138

Db         66 PPERKPWGDQYDYLNTRP-VELV--SKKGTDGVPVMLQTNFFRLKTKPEWRIVHYHVEFE 122

Qy        139 PLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVTEVFSQTRNGEHVRITITL 198

Db        123 PSIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIKF 182

Qy        199 TNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSIL 258

Db        183 VGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSIR 242

Qy        259 QYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTY 318

Db        243 QHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRTY 300

Qy        319 RVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLP 378

Db        301 RINDVDFGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNAS 358

Qy        379 GPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQ-- 436

Db        359 ELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPES 415

Qy        437 -RELRDWGLSFDSNLLSFSGRILQSEKI--HQGGKTFDYNPQFADWSKETRGAPLISVKP 493

Db        416 VKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTPS 472

Qy        494 --LDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKK-AIMIEVDDRTEAYLRALQQKV 550

Db        473 DGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCV 532

Qy        551 TSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMN 610

Db        533 RSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKKPYSLMSIATKIAIQLN 592

Qy        611 CKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASIN-EGMTRWFSRCVFQDRG 669

Db        593 CKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAF 652

Qy        670 QELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVG 729

Db        653 DVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEY 712

Qy        730 RGYN---PRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVR 786

Db        713 ARVQLSPPQLAYIVVTRSMNTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVSQQVR 768

Qy        787 SGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVGQS 846

Db        769 QGTVSPTSYNVLYSSMGLSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTN 828

Qy        847 IHREPNLSLSNRLYYL 862

Db        829 LHSIPQNALEKKFYYL 844
```

RESULT 8
US-10-029-495-7
; Sequence 7, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
; APPLICANT: Drmanac, Radoje T.
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Dickson, Mark C.
; APPLICANT: Jones, Lee William
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Xue, Aidong
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi,Vinod
; TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
; FILE REFERENCE: 28110/35905A
; CURRENT APPLICATION NUMBER: US/10/029,495
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 09/378,667
; PRIOR FILING DATE: 1999-08-20
; PRIOR APPLICATION NUMBER: 09/687,527

```
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 844
;   TYPE: PRT
;   ORGANISM: Drosophila
;   FEATURE:
;   OTHER INFORMATION: piwi
US-10-029-495-7
```

Query Match 31.0%; Score 1404; DB 15; Length 844;
Best Local Similarity 35.9%; Pred. No. 7e-129;
Matches 307; Conservative 171; Mismatches 320; Indels 58; Gaps 17;

```
Qy         48 GRGRQRGMVVGATSKSQ----------------------ELQISAGFQELSLAE----RGG 82
Db          6 GRGRRRPLNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG 65

Qy         83 --RRRDFHD--XGVNTRQNLDHVKESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYN 138
Db         66 PTERKPWGDQYDYLNTR---PAELVSKKGTDGVPVMLQTNFFRLKTKPEWRIVHYHVEFE 122

Qy        139 PLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVTEVFSQTRNGEHVRITITL 198
Db        123 PSIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIKF 182

Qy        199 TNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSIL 258
Db        183 VGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSIR 242

Qy        259 QYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTY 318
Db        243 QHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRTY 300

Qy        319 RVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLP 378
Db        301 RINDVDFGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNAS 358

Qy        379 GPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQ-- 436
Db        359 ELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPES 415

Qy        437 -RELRDWGLSFDSNLLSFSGRILQSEKI--HQGGKTFDYNPQFADWSKETRGAPLISVKP 493
Db        416 VKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTPS 472

Qy        494 --LDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKK-AIMIEVDDRTEAYLRALQQKV 550
Db        473 DGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCV 532

Qy        551 TSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMN 610
Db        533 RSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKKPYSLMSIATKIAIQLN 592

Qy        611 CKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASIN-EGMTRWFSRCVFQDRG 669
Db        593 CKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAF 652

Qy        670 QELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVG 729
Db        653 DVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEY 712

Qy        730 RGYN---PRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVR 786
Db        713 ARVQLSPPQLAYIVVTRSMNTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVSQQVR 768

Qy        787 SGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVGQS 846
Db        769 QGTVSPTSYNVLYSSMGLSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTN 828

Qy        847 IHREPNLSLSNRLYYL 862
```

```
Db        829 LHSIPQNALEKKFYYL 844

RESULT 9
US-09-873-737A-2
; Sequence 2, Application US/09873737A
; Patent No. US20020076797A1
; GENERAL INFORMATION:
;  APPLICANT: Duke University
;  APPLICANT:  Lin, Haifan
;  TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
;  TITLE OF INVENTION:  PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
;  FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2
;  CURRENT APPLICATION NUMBER: US/09/873,737A
;  CURRENT FILING DATE:  2001-06-04
;  PRIOR APPLICATION NUMBER: PCT/US99/28764
;  PRIOR FILING DATE: 1999-12-03
;  PRIOR APPLICATION NUMBER: 60/110,901
;  PRIOR FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 843
;   TYPE: PRT
;   ORGANISM: Drosophila sp.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (13)
;   OTHER INFORMATION: Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (106)
;   OTHER INFORMATION: Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (785)
;   OTHER INFORMATION: Xaa=Leu or Ile
US-09-873-737A-2
```

Query Match 30.6%; Score 1385.5; DB 10; Length 843;
Best Local Similarity 35.7%; Pred. No. 4.7e-127;
Matches 306; Conservative 173; Mismatches 318; Indels 59; Gaps 19;

```
Qy         48 GRGRQRGMVVGATSKSQ----------------------ELQISAGFQELSLAE----RGG 82
Db          6 GRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG 65

Qy         83 --RRRDFHD--XGVNTRQNLDHVKESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYN 138
Db         66 PPERKPWGDQYDYLNTRP-VELV--SKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEFV 122

Qy        139 PLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVTEVFSQTRNGEHVRITITL 198
Db        123 PTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIKF 182

Qy        199 TNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSIL 258
Db        183 VGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSIR 242

Qy        259 QYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTY 318
Db        243 QHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRTY 300

Qy        319 RVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLP 378
Db        301 RINDVDFGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNAS 358

Qy        379 GPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQ-- 436
Db        359 ELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPES 415

Qy        437 -RELRDWGLSFDSNLLSFSGRILQSEKI--HQGGKTFDYNPQFADWSKETRGAPLISVKP 493
Db        416 VKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTPS 472
```

```
Qy        494 --LDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKK-AIMIEVDDRTEAYLRALQQKV 550
                 || | :|  :||     :|: :|::    ||::::     |  ||||  |:||:   |
Db        473 DGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCV 532

Qy        551 TSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMN 610
               || :::::||: ::  ::| :||| |   |:| |  :|  | :::|:|||||:|:|
Db        533 RSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQLN 591

Qy        611 CKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASIN-EGMTRWFSRCVFQDRG 669
               ||:|   | ::: |   | :| | |   :|:   :||:: :  : :||
Db        592 CKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAF 651

Qy        670 QELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVG 729
               | : |   :  ||| :   :  :|||:: |||||  | || |  :||   :: ||:
Db        652 DVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEY 711

Qy        730 RGYN---PRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVR 786
                   |:|  ||| : :| |||       ||| |||::|  :| || |||::||| ||
Db        712 ARVQLSPPQLAYIVVTRSMNTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVSQQVR 767

Qy        787 SGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVGQS 846
               |:|||| |||:| | |  |: :|:|||| ||:|||| |  |||| |||| ||| ||| :
Db        768 QGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTN 827

Qy        847 IHREPNLSLSNRLYYL 862
               :|  |  :|  : |||
Db        828 LHSIPQNALEKKFYYL 843
```

RESULT 10
US-09-801-574-16
; Sequence 16, Application US/09801574
; Patent No. US20020081592A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Peijing Jeremy
; APPLICANT: Page, David C.
; TITLE OF INVENTION: Reproduction-Specific Genes
; FILE REFERENCE: 0399.2007-002
; CURRENT APPLICATION NUMBER: US/09/801,574
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: 60/187,518
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: 60/261,557
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 580
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-801-574-16

Query Match 30.0%; Score 1360.5; DB 10; Length 580;
Best Local Similarity 43.3%; Pred. No. 7.2e-125;
Matches 253; Conservative 121; Mismatches 201; Indels 9; Gaps 8;

```
Qy        284 MFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSF 343
               |  :||| :|| ||:: || |:| ||:|:|||:|||:||:|:: || :|  :|| |::|
Db          1 MHAIYQQNKEH-FQDECSKLLVGSIVITRYNNRTYRIDDVDWNKTPKDSFVMSDGKEITF 59

Qy        344 LEYYRKQYNQEITDLKQPVLVSQP-KRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFN 402
               |||| | |   : :  ||:|: :| :|:   |  | |  : :||| :|||: :||: ||
Db         60 LEYYSKNYGITVKEDDQPLLIHRPSERQNNHGMLLKGEILLLPELSFMTGIPEKMKKDFR 119

Qy        403 VMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEK 462
               ||||     |:|:|    :  |:  | :::    || ||||   ::    ||:|  |:
Db        120 AMKDLTQQINLSPKQHHGALECLLQRISQNETASNELTRWGLSLHKDVHKIEGRLLPMER 179

Qy        463 IHQGGKTFDYNPQFADWSKE-TRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTP 521
               |:   :|   :  :| || || | :::: |:  | | | :|  : |  |:  | |:
Db        180 INLRNTSF-VTSEGLNWVKEVTRDASILTI-PMHFWALFYPKRAMDQARELVNMLEKIAG 237

Qy        522 AMGIQMKKAIMIEV-DDRTEAYLRALQQ--KVTSDTQIVVCLLSSNRKDKYDAIKKYLCT 578
```

```
Db        238 PIGMRTSPPAWVELKDDRIETYIRTIQSLLGVEGKIQMVVCIIMGTRDDLYGAIKKLCCV 297

Qy        579 DCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDT 638
               | ||| :  ||:|:   : ::| || ||||||:||||| ||: ||  |::|:| |||
Db        298 QSPVPSQVINVRTIGQPTRLRSVAQKILLQMNCKLGGELWGVDIPLKQLMVIGMDVYHDP 357

Qy        639 TAGRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRV 698
               : | ||: |||||||  :|:|:|| |||   ||:|| ||:|| :|: :   |  :| ::
Db        358 SRGMRSVVGFVASINLTLTKWYSRVVFQMPHQEIVDSLKLCLVGSLKKYYEVNHCLPEKI 417

Qy        699 IVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQ 758
               :|||||| ||||||: |||:||   | ::    |:|:: | ||:|:::   :  :
Db        418 VVYRDGVSDGQLKTVANYEIPQLQKCFEAFD-NYHPKMVVFVVQKKISTNLYLAAPDHFV 476

Qy        759 NPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCH 818
               | ||||:|  :|  || ||:::  || |   ||||  : ::: | |||:|||||:| ||
Db        477 TPSPGTVVDHTITSCEWVDFYLLAHHVRQGCGIPTHYICVLNTANLSPDHMQRLTFKLCH 536

Qy        819 VYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 862
               :|:|||| |||||||:||||||| || :| || : |   |::|
Db        537 MYWNWPGTIRVPAPCKYAHKLAFLSGQILHHEPAIQLCGNLFFL 580
```

RESULT 11
US-10-029-495-9
; Sequence 9, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
; APPLICANT: Drmanac, Radoje T.
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Dickson, Mark C.
; APPLICANT: Jones, Lee William
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Xue, Aidong
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi,Vinod
; TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
; FILE REFERENCE: 28110/35905A
; CURRENT APPLICATION NUMBER: US/10/029,495
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 09/378,667
; PRIOR FILING DATE: 1999-08-20
; PRIOR APPLICATION NUMBER: 09/687,527
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-029-495-9

Query Match 28.7%; Score 1299.5; DB 15; Length 498;
Best Local Similarity 50.4%; Pred. No. 5.8e-119;
Matches 254; Conservative 87; Mismatches 154; Indels 9; Gaps 4;

```
Qy        362 VLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQRE 421
               :|||  |::|      |  |  |||||:||||||:  :|| :|| :|  |||:|  ||:
Db          1 MLVSLLKKKRNDNSE-PQLAHLIPELCFLTGLTDQATSDFQLMKAVAEKTRLSPSGRQQR 59

Qy        422 VGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQ---FAD 478
               : ||:| | :: | : ||  ||| | |  :| :|||: ||||       |:  |    ||
Db         60 LARLVDNIQRNTNARFELETWGLHFGSQ-ISLTGRIVPSEKILMQ----DHICQPVSAAD 114

Qy        479 WSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEVDDR 538
               |||: |   ::: : |: ||:: : |     | | :  | :|  :||  :    :|:| :
Db        115 WSKDIRTCKILNAQSLNTWLILCSDRTEYVAESFLNCLRRVAGSMGFNVDYPKIIKVQEN 174

Qy        539 TEAYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTV 598
               |::||:|| |  | |:|:|:| ||:|  ||:|||||  :||| |||||:|||| ||  :
```

```
Db        175 PAAFVRAIQQYVDPDVQLVMCILPSNQKTYYDSIKKYLSSDCPVPSQCVLARTLNKQGMM 234

Qy        599 MAIATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTR 658

Db        235 MSIATKIAMQMTCKLGGELWAVEIPLKSLMVVGIDVCKDALSKDVMVVGCVASVNPRITR 294

Qy        659 WFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEV 718

Db        295 WFSRCILQRTMTDVADCLKVFMTGALNKWYKYNHDLPARIIVYRAGVGDGQLKTLIEYEV 354

Qy        719 PQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDF 778

Db        355 PQLLSSVAESSSNTSSRLSVIVVRKKCMPRFFTEMNRTVQNPPLGTVVDSEATRNEWYDF 414

Qy        779 FIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHK 838

Db        415 YLISQVACRGTVSPTYYNVIYDDNGLKPDHMQRLTFKLCHLYYNWPGIVSVPAPCQYAHK 474

Qy        839 LAFLVGQSIHREPNLSLSNRLYYL 862

Db        475 LTFLGAQSIHKEPSLELANHLFYL 498


RESULT 12
US-10-029-495-6
; Sequence 6, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
;  APPLICANT: Drmanac, Radoje T.
;  APPLICANT:  Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Dickson, Mark C.
;  APPLICANT:  Jones, Lee William
;  APPLICANT:  Ballinger, Dennis G.
;  APPLICANT:  Xue, Aidong
;  APPLICANT:  Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi,Vinod
;  TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
;  FILE REFERENCE: 28110/35905A
;  CURRENT APPLICATION NUMBER: US/10/029,495
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: 09/378,667
;  PRIOR FILING DATE: 1999-08-20
;  PRIOR APPLICATION NUMBER: 09/687,527
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 824
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
;   FEATURE:
;   OTHER INFORMATION: ciwi2
US-10-029-495-6

  Query Match           28.5%;  Score 1290;  DB 15;  Length 824;
  Best Local Similarity   33.7%;  Pred. No. 1.2e-117;
  Matches  289;  Conservative  176;  Mismatches  311;  Indels   82;  Gaps   19;

Qy         48 GRGRQRGMVVGATSKSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTG 107

Db          6 GRGRGRGSGSNNSGGKDQKYLGTIQPDLFIRQQG----------------------QSKTG 44

Qy        108 SSGIIVKLSTNHFRL-TSRPQWALYQYHIDYNPLMEARRLRSALLFQHE--DLIGRCHAF 164

Db         45 SSGQPQKCFANFIPIEMTQSDYSIYQYHVEFEPTVDSKANREKMLRDNNVTDEIGHHFVF 104

Qy        165 DGTILFLPKRL-QHKVTEVFSQTRNGEH------VRITITLTNELPPTSPTCLQFYNIXF 217

Db        105 DGMILYLKEEWEQNQMIEV-------QHPIDRSLICIRFKQTNRFLVDDPQTINIFNTII 157

Qy        218 RRLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRS 277

Db        158 RRSFDALQLTQLGRNYFNWGDSRAVPDYNMSILPGYETAIRMYEENFMLCVENRFKMVRE 217

Qy        278 ETVLDFMFNLYQ------QTEEHKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKS 331

Db        218 ES----MYILFHKELRSCQNNPQRVQEKMNEMYGGTTIITRYNNKLHRYTRLDYSISPLS 273

Qy        332 TFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMXIPELCYLT 391

Db        274 EFVK-DGQSITLKEYFKNQYGIEITVDDQPIIISEGKPKQ--PGEPPQVSYIVPELCFPT 330

Qy        392 GLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLL 451

Db        331 GLTDEMRKDFKMMKEIAKHTRMSPQQRLVESRKLIVDLSKNEKVMECFKYWGISLGQDLA 390

Qy        452 SFSGRILQSEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEA--A 509

Db        391 NVQARVLKSEPL-QGKKT--YEGKQAEWARGVKECGIYRGSNMTNWIVIGPGSGNSGLLS 447

Qy        510 NSLIQNLFKVTPAMGIQMKKAIMIEV-----DDRTEAYLRALQQKVTSDTQIVVCLLSSN 564

Db        448 QKFIEEARRLGKILQVQLGEPMCVPIRGISPNDYLEGVKGAIKQVDGEDIHMLVVMLADD 507

Qy        565 RKDKYDAIKKYLCTDCPTPSQCVVARTL-------GKQQTVMAIATKIALQMNCKMGGEL 617

Db        508 NKTRYDSLKKFLCVECPIPNQCVNLRTLAGKSKDGGENKNLGSIVLKIVLQMICKTGGAL 567

Qy        618 WRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLK 677

Db        568 WKVNIPLKNTMIVGYDLYHDSTLKGKTVGACVSTTSNDFTQFYSQTRPHENPTQLGNNLT 627

Qy        678 VCLQAALRAWSGCNEY-MPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGR------ 730

Db        628 HFVRKALKQYYDSNDQTLPSRLILYRDGAGDGQIPYIKNTEVKLVRDACDAVTDKAAELS 687

Qy        731 ---GYNPRLTVIVVKKRVNARFFAQSGGRLQ---NPLPGTVIDVEVTRPEWYDFFIVSQA 784

Db        688 NKVQEKIKLAFIIVTKRVNMRILKQ-GSSLDNAINPQPGTVVDTTVTRPERMDFYLVPQF 746

Qy        785 VRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVG 844

Db        747 VNQGTVTPVSYNIIHDDTDLGPDKHQQLAFKLCHLYYNWQGTVRVPAPCQYAHKLAFLTA 806

Qy        845 QSIHREPNLSLSNRLYYL 862

Db        807 QSLHDDANGCLRDKLFFL 824


RESULT 13
US-10-029-495-5
; Sequence 5, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
;  APPLICANT: Drmanac, Radoje T.
;  APPLICANT:  Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Dickson, Mark C.
;  APPLICANT:  Jones, Lee William
;  APPLICANT:  Ballinger, Dennis G.
;  APPLICANT:  Xue, Aidong
;  APPLICANT:  Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi,Vinod
;  TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
;  FILE REFERENCE: 28110/35905A
;  CURRENT APPLICATION NUMBER: US/10/029,495
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: 09/378,667
;  PRIOR FILING DATE: 1999-08-20
;  PRIOR APPLICATION NUMBER: 09/687,527
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
```

; SEQ ID NO 5
; LENGTH: 722
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:
; OTHER INFORMATION: ciwi1
US-10-029-495-5

Query Match 25.1%; Score 1137; DB 15; Length 722;
Best Local Similarity 33.4%; Pred. No. 1.2e-102;
Matches 255; Conservative 158; Mismatches 275; Indels 76; Gaps 17;

```
Qy        128 WALYQYHIDYNPLMEARRLRSALLFQHEDL--IGRCHAFDGTILFLPKRL-QHKVTEVFS 184
Db          6 YSIYQYHVEFEPTVDSKATRENMLRQPSVTVEIGKHFVFDGMILYLKEEWDQNQMIEV-Q 64

Qy        185 QTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIPN 244
Db         65 HPNDNSLICIRFKKTNRFLVDDPQTINIFNTIIRRSFDAMKLTQIGRNYFD--------- 115

Qy        245 HRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQVSKEL 304
Db        116 --------------WDNSRAL-----RKELRS----------CQNNRQRVQEKMNEVY 144

Qy        305 IGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLV 364
Db        145 GGSTIITRYNNKLHRFTRLDNEITPLSKFQK-DGEQIILKEYFKNQYDIDITDDEQFIII 203

Qy        365 SQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGR 424
Db        204 SEGKPKQ--PGEPPQVNYIVPELCFPTGLTDEMRKDFKMMKEIAKHTRMSPQQRLDETRK 261

Qy        425 LIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWSKETR 484
Db        262 LITKLSQNQTMMECFQYWGISLGQDLANVQARVLKSEPL-QGKR--QYEGKQAEWARGVK 318

Qy        485 GAPLISVKPLDNWLLIYTRRNYEA--ANSLIQNLFKVTPAMGIQMKKAIMIEVDD-RTEA 541
Db        319 ECGIYRGSNMTNWIVIGPGSGNSGLLAQKFIAEARNLGRTLQVQLGEPMCVKINGISPND 378

Qy        542 YLRALQQKVTS----DTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTL----- 592
Db        379 YLEGLKAAIKSVDGEEIHMLVVMLADDNKTRYDSLKKYLCVECPIPNQCVNLRTLAGKSK 438

Qy        593 --GKQQTVMAIATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVA 650
Db        439 DGGENKNLGSIVLKIVLQMICKTGGALWKVNIPLKSTMIVGYDLYHDSTLKGKTVGACVS 498

Qy        651 SINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNE-YMPSRVIVYRDGVGDGQ 709
Db        499 TTSNDFTQFYSQTRPHENPTQLGNNLTHFVRKSLKQYYDNNDKTLPSRLILYRDGAGDGQ 558

Qy        710 LKTLVNYEVPQFLDCLKSVGR---------GYNPRLTVIVVKKRVNARFFAQ--SGGRLQ 758
Db        559 IPYIKNTEVKLVRDACDAVTDKAAELSNKVQEKIKLAFIIVTKRVNMRILKQGSSSKSAI 618

Qy        759 NPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCH 818
Db        619 NPQPGTVVDTTVTRPERMDFYLVPQFVNQGTVTPVSYNIIHDDTGLGPDKHQQLAFKLCH 678

Qy        819 VYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 862
Db        679 LYYNWQGTVRVPAPCQYAHKLAFLTAQSLHDDANGYLRDKLFFL 722
```

RESULT 14
US-10-102-806-685
; Sequence 685, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103P1C1
; CURRENT APPLICATION NUMBER: US/10/102,806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 685
; LENGTH: 130
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (61)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-102-806-685

Query Match 13.7%; Score 620; DB 15; Length 130;
Best Local Similarity 92.7%; Pred. No. 6.1e-53;
Matches 115; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

```
Qy        579 DCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDT 638
Db          5 DCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCXHDM 64

Qy        639 TAGRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRV 698
Db         65 TAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRI 124

Qy        699 IVYR 702
Db        125 IVYR 128
```

RESULT 15
US-10-174-363-54
; Sequence 54, Application US/10174363
; Publication No. US20030077623A1
; GENERAL INFORMATION:
; APPLICANT: Butler, Karlene H.
; APPLICANT: Cahoon, Rebecca E
; APPLICANT: Harvell, Leslie T.
; APPLICANT: Rafalski, Antoni J.
; APPLICANT: Sakai, Hajime
; TITLE OF INVENTION: Polynucleotides And Polypeptides Involved In Post-Transcriptio
; TITLE OF INVENTION: Gene Silencing
; FILE REFERENCE: BB1454 US NA
; CURRENT APPLICATION NUMBER: US/10/174,363
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: 60/298,973
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 54
; LENGTH: 1101
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-174-363-54

Query Match 11.3%; Score 511; DB 15; Length 1101;
Best Local Similarity 22.7%; Pred. No. 1.3e-40;
Matches 230; Conservative 156; Mismatches 437; Indels 188; Gaps 38;

```
Qy          3 GRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEGDLVGRG--RQRG------ 54
Db         83 GRGGPGSHHPGGGPPDYQGRGGSGSHHPGGGPPEYQPR----DYQGRGGPRPRGGMPQPY 138

Qy         55 ---------------------------------MV----VGATSKSQ-ELQISAG---- 71
Db        139 YGGPRGSGGRSVPSGSSRTVPELHQAPHVQYQAPMVSPTPSGAGSSSQPAAEVSSGQVQQ 198
```

```
Qy         72 -FQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHFRLTSRPQWAL 130
Db        199 QFQQLATRDQSSTSQAIQIAPPSSKSVRFPLRPGK-GTYGDRCIVKANHF-FAELPDKDL 256

Qy        131 YQYHIDYNPLMEARRLRSALLFQ------HEDLIGRCHAFDG-----TILFLP------- 172
Db        257 HQYDVSITPEVTSRGVNRAVMFELVTLYRYSHLGGRLPAYDGRKSLYTAGPLPFASRTFE 316

Qy        173 ----------------KRLQHKVTEVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIX 216
Db        317 ITLQDEEDSLGGGQGTQRRERLFRVVIKFAARADLHHLAMFLAGRQADAPQEALQVLDIV 376

Qy        217 FRRLLKIMNLQQIGRNYYNPSDPIDIPNHR-----LVIWPGFTTSILQYENNIMLCTDVS 271
Db        377 LRE-LPTTRYSPVGRSFYSP----NLGRRQQLGEGLESWRGFYQSIRPTQMGLSLNIDMS 431

Qy        272 HKV-LRSETVLDFMFNLYQ-----QTEEHKFQEQVSKELIGL-IVLTKYNN--KTYRVDD 322
Db        432 STAFIEPLPVIDFVAQLLNRDISVRPLSDSDRVKIKKALRGVKVEVTHRGNMRRKYRISG 491

Qy        323 IDWDQNPKSTFKKAD-GSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPA 381
Db        492 LTSQATRELSFPVDDRGTVKTVVQYFLETYGFSIQHTTLPCLQVGNQQR---------PN 542

Qy        382 MXIPELCYLT--GLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDY--IHKDDNVQR 437
Db        543 YLPMEVCKIVEGQRYSKRLNEKQITALLKV-TCQRPQERELDILRTVSHNAYHEDQYAQ- 600

Qy        438 ELRDWGLSFDSNLLSFSGRILQSE--KIHQGGKTFDYNPQFADWSKETRGAPLISVKPLD 495
Db        601 ---EFGIKIDERLASVEARVLPPPRLKYHDSGREKDVLPRVGQWNMMNK--KMVNGGRVN 655

Qy        496 NWLLIYTRRNYE--AANSLIQNLFKVTPAMGIQMK-KAIMIEVDDRTEAYLRALQQKVTS 552
Db        656 NWACINFSRNVQDSAARGFCHELAIMCQISGMDFALEPVLPPLTARPEHVERALKARYQD 715

Qy        553 ----------DTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIA 602
Db        716 AMNMLRPQGRELDLLIVILPDNNGSLYGDLKRICETDLGLVSQCCLTKHVFKMS--KQYL 773

Qy        603 TKIALQMNCKMGGELWRVDMAL---------KLAMIVGIDCYH-----DTTAGRRSIAGF 648
Db        774 ANVALKINVKVGGRNTVLVDALTRRIPLVSDRPTIIFGADVTHPHPGEDSSP---SIAAV 830

Qy        649 VASIN-EGMTRWFSRCVFQDRGQELVDGL-KV------------CLQAALRAWSGCNEYM 694
Db        831 VASQDWPEVTKYAGLVSAQAHRQELIQDLFKVWQDPHRGTVTGGMIKELLISFKRATGQK 890

Qy        695 PSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSG 754
Db        891 PQRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEPNYQPPVTFVVVQKRHHTRLFANNH 950

Qy        755 G------RLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDH 808
Db        951 NDQRTVDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDENKFTADE 1010

Qy        809 IQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRL 859
Db       1011 LQTLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRA--RFYMEPETSDSGSM 1059
```

Search completed: July 22, 2003, 07:07:51
Job time : 24.1715 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 06:25:13 ; Search time 17.1325 Seconds
(without alignments)
4836.881 Million cell updates/sec

Title: US-09-873-737A-4
Perfect score: 4528
Sequence: 1 MTGRARARARGRARGQETVQ.....VGQSIHREPNLSLSNRLYYL 862

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1290	28.5	824	2	T20351	hypothetical prote
2	1137	25.1	722	2	T30995	hypothetical prote
3	541.5	12.0	834	2	T41568	argonaute-like pro
4	522.5	11.5	1014	2	H86438	protein T19E23.7 [
5	495.5	10.9	988	2	T52134	Zwille protein [im
6	489.5	10.8	891	2	T32079	hypothetical prote
7	489.5	10.8	1000	2	T22391	hypothetical prote
8	462.5	10.2	930	2	A84668	Argonaute (AGO1)-l
9	460	10.2	813	2	JC6569	translation initia
10	460	10.2	997	2	A84678	Argonaute (AGO1)-l
11	428.5	9.5	1032	2	T23164	hypothetical prote
12	428.5	9.5	1035	2	T23165	hypothetical prote
13	423	9.3	958	2	S41013	hypothetical prote
14	423	9.3	1040	2	D88568	protein ZK757.3 [i
15	417.5	9.2	887	2	T01113	translation initia
16	381.5	8.4	898	2	T25164	hypothetical prote
17	326.5	7.2	1139	2	T33275	hypothetical prote
18	321.5	7.1	905	2	T23510	hypothetical prote
19	290.5	6.4	892	2	T15195	hypothetical prote
20	283	6.2	945	2	T23965	hypothetical prote
21	247	5.5	965	2	T22933	hypothetical prote
22	237	5.2	1032	2	T19324	hypothetical prote
23	237	5.2	1057	2	T16676	hypothetical prote
24	233.5	5.2	990	2	F88925	protein T22H9.3 [i
25	226.5	5.0	697	2	T15179	hypothetical prote
26	223	4.9	939	2	T18974	hypothetical prote
27	220	4.9	1030	2	T16114	hypothetical prote
28	218	4.8	363	2	T34486	hypothetical prote
29	211	4.7	324	2	G87774	protein C24A11.3 [
30	211	4.7	674	2	T19268	hypothetical prote
31	210	4.6	881	2	T31818	hypothetical prote
32	194.5	4.3	706	2	H70425	conserved hypothet
33	188	4.2	567	2	T34339	hypothetical prote
34	165	3.6	944	2	T18911	hypothetical prote
35	133.5	2.9	709	2	G96610	probable disease r
36	127	2.8	713	2	H64464	hypothetical prote
37	122.5	2.7	185	2	T27784	hypothetical prote
38	122.5	2.7	1238	2	S63247	probable membrane
39	121	2.7	949	2	G84486	probable plasma me
40	118.5	2.6	1527	1	RNZMB2	DNA-directed RNA p
41	115.5	2.6	1766	2	S03701	141K protein - pea
42	113.5	2.5	690	2	H82923	conserved hypothet
43	113.5	2.5	1547	2	JQ0096	hypothetical 176K
44	113	2.5	930	2	T08588	hypothetical prote
45	113	2.5	1120	2	H71664	transcription-repa

ALIGNMENTS

RESULT 1
T20351
hypothetical protein D2030.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T20351
R;Wilkinson, J.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19261
A;Accession: T20351
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-824 <WIL>
A;Cross-references: EMBL:Z73906; PIDN:CAA98113.1; GSPDB:GN00019; CESP:D2030.6
A;Experimental source: clone D2030
C;Genetics:
A;Gene: CESP:D2030.6
A;Map position: 1
A;Introns: 118/3; 223/3; 304/3; 452/1; 573/3; 632/1; 711/2; 773/3

Query Match 28.5%; Score 1290; DB 2; Length 824;
Best Local Similarity 33.7%; Pred. No. 1.2e-88;
Matches 289; Conservative 176; Mismatches 311; Indels 82; Gaps 19;

```
Qy         48 GRGRQRGMVVGATSKSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTG 107
              |||| ||     :    : :     :| : ::|                      :||||
Db          6 GRGRGRGSGSNNSGGKDQKYLGTIQPDLFIRQQG--------------------QSKTG 44

Qy        108 SSGIIVKLSTNHFRL-TSRPQWALYQYHIDYNPLMEARRLRSALLFQHE--DLIGRCHAF 164
              |||   |   |  :  ::  :::||||::: | :::: |  :|  :   | ||    |
Db         45 SSGQPQKCFANFIPIEMTQSDYSIYQYHVEFEPTVDSKANREKMLRDNNVTDEIGHHFVF 104

Qy        165 DGTILFLPKRL-QHKVTEVFSQTRNGEH------VRITITLTNELPPTSPTCLQFYNIXF 217
              || ||:| :   |::: ||       :|       : |    ||        | : :|
Db        105 DGMILYLKEEWEQNQMIEV-------QHPIDRSLICIRFKQTNRFLVDDPQTINIFNTII 157

Qy        218 RRLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRS 277
              ||    : | |:|||||:|  |   :|:: : | ||: |:|  || | ||| :   |::|
Db        158 RRSFDALQLTQLGRNYFNWGDSRAVPDYNMSILPGYETAIRMYEENFMLCVENRFKMVRE 217

Qy        278 ETVLDFMFNLYQ------QTEEHKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKS 331
              |:    |: |:  *    |    : ||:::   |  ::|:|||| :|   :|:  :| |
Db        218 ES----MYILFHKELRSCQNNPQRVQEKMNEMYGGTTIITRYNNKLHRYTRLDYSISPLS 273

Qy        332 TFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMXIPELCYLT 391
              | | ||  ::  ||:: ||  |||   ||:::|: | ::   |  |  :  :||||: |
Db        274 EFVK-DGQSITLKEYFKNQYGIEITVDDQPIIISEGKPKQ--PGEPPQVSYIVPELCFPT 330

Qy        392 GLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLL 451
              ||||:|| || :||::| |||::|:||  |  :||  : |:: |     : ||:|   :|
Db        331 GLTDEMRKDFKMMKEIAKHTRMSPQQRLVESRKLIVDLSKNEKVMECFKYWGISLGQDLA 390
```

```
Qy        452 SFSGRILQSEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEA--A 509
Db        391 NVQARVLKSEPL-QGKKT--YEGKQAEWARGVKECGIYRGSNMTNWIVIGPGSGNSGLLS 447

Qy        510 NSLIQNLFKVTPAMGIQMKKAIMIEV-----DDRTEAYLRALQQKVTSDTQIVVCLLSSN 564
Db        448 QKFIEEARRLGKILQVQLGEPMCVPIRGISPNDYLEGVKGAIKQVDGEDIHMLVVMLADD 507

Qy        565 RKDKYDAIKKYLCTDCPTPSQCVVARTL-------GKQQTVMAIATKIALQMNCKMGGEL 617
Db        508 NKTRYDSLKKFLCVECPIPNQCVNLRTLAGKSKDGGENKNLGSIVLKIVLQMICKTGGAL 567

Qy        618 WRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLK 677
Db        568 WKVNIPLKNTMIVGYDLYHDSTLKGKTVGACVSTTSNDFTQFYSQTRPHENPTQLGNNLT 627

Qy        678 VCLQAALRAWSGCNEY-MPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGR------ 730
Db        628 HFVRKALKQYYDSNDQTLPSRLILYRDGAGDGQIPYIKNTEVKLVRDACDAVTDKAAELS 687

Qy        731 ---GYNPRLTVIVVKKRVNARFFAQSGGRLQ---NPLPGTVIDVEVTRPEWYDFFIVSQA 784
Db        688 NKVQEKIKLAFIIVTKRVNMRILKQ-GSSLDNAINPQPGTVVDTTVTRPERMDFYLVPQF 746

Qy        785 VRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVG 844
Db        747 VNQGTVTPVSYNIIHDDTDLGPDKHQQLAFKLCHLYYNWQGTVRVPAPCQYAHKLAFLTA 806

Qy        845 QSIHREPNLSLSNRLYYL 862
Db        807 QSLHDDANGCLRDKLFFL 824
```

RESULT 2
T30995
hypothetical protein C01G5.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 29-Oct-1999
C;Accession: T30995
R;Bradshaw, H.; Stellyes, L.
submitted to the EMBL Data Library, August 1999
A;Description: The sequence of C. elegans cosmid C01G5.
A;Reference number: Z20956
A;Accession: T30995
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-722 <BRA>
A;Cross-references: EMBL:U50068; PIDN:AAB37734.1
A;Experimental source: strain Bristol N2
C;Genetics:
A;Map position: IV
A;Introns: 58/3; 121/3; 202/3; 350/1; 471/3; 530/1; 609/2; 671/3
A;Note: C01G5.2

 Query Match 25.1%; Score 1137; DB 2; Length 722;
 Best Local Similarity 33.4%; Pred. No. 3.3e-77;
 Matches 255; Conservative 158; Mismatches 275; Indels 76; Gaps 17;

```
Qy        128 WALYQYHIDYNPLMEARRLRSALLFQHEDL--IGRCHAFDGTILFLPKRL-QHKVTEVFS 184
Db          6 YSIYQYHVEFEPTVDSKATRENMLRQPSVTVEIGKHFVFDGMILYLKEEWDQNQMIEV-Q 64

Qy        185 QTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIPN 244
Db         65 HPNDNSLICIRFKKTNRFLVDDPQTINIFNTIIRRSFDAMKLTQIGRNYFD--------- 115

Qy        245 HRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQVSKEL 304
Db        116 --------------WDNSRAL-----RKELRS----------CQNNRQRVQEKMNEVY 144

Qy        305 IGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLV 364
Db        145 GGSTIITRYNNKLHRFTRLDNEITPLSKFQK-DGEQIILKEYFKNQYDIDITDDEQFIII 203

Qy        365 SQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGR 424
Db        204 SEGKPKQ--PGEPPQVNYIVPELCFPTGLTDEMRKDFKMMKEIAKHTRMSPQQRLDETRK 261

Qy        425 LIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWSKETR 484
Db        262 LITKLSQNQTMMECFQYWGISLGQDLANVQARVLKSEPL-QGKR--QYEGKQAEWARGVK 318

Qy        485 GAPLISVKPLDNWLLIYTRRNYEA--ANSLIQNLFKVTPAMGIQMKKAIMIEVDD-RTEA 541
Db        319 ECGIYRGSNMTNWIVIGPGSGNSGLLAQKFIAEARNLGRTLQVQLGEPMCVKINGISPND 378

Qy        542 YLRALQQKVTS----DTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTL----- 592
Db        379 YLEGLKAAIKSVDGEEIHMLVVMLADDNKTRYDSLKKYLCVECPIPNQCVNLRTLAGKSK 438

Qy        593 --GKQQTVMAIATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVA 650
Db        439 DGGENKNLGSIVLKIVLQMICKTGGALWKVNIPLKSTMIVGYDLYHDSTLKGKTVGACVS 498

Qy        651 SINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNE-YMPSRVIVYRDGVGDGQ 709
Db        499 TTSNDFTQFYSQTRPHENPTQLGNNLTHFVRKSLKQYYDNNDKTLPSRLILYRDGAGDGQ 558

Qy        710 LKTLVNYEVPQFLDCLKSVGR---------GYNPRLTVIVVKKRVNARFFAQ--SGGRLQ 758
Db        559 IPYIKNTEVKLVRDACDAVTDKAAELSNKVQEKIKLAFIIVTKRVNMRILKQGSSSKSAI 618

Qy        759 NPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCH 818
Db        619 NPQPGTVVDTTVTRPERMDFYLVPQFVNQGTVTPVSYNIIHDDTGLGPDKHQQLAFKLCH 678

Qy        819 VYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 862
Db        679 LYYNWQGTVRVPAPCQYAHKLAFLTAQSLHDDANGYLRDKLFFL 722
```

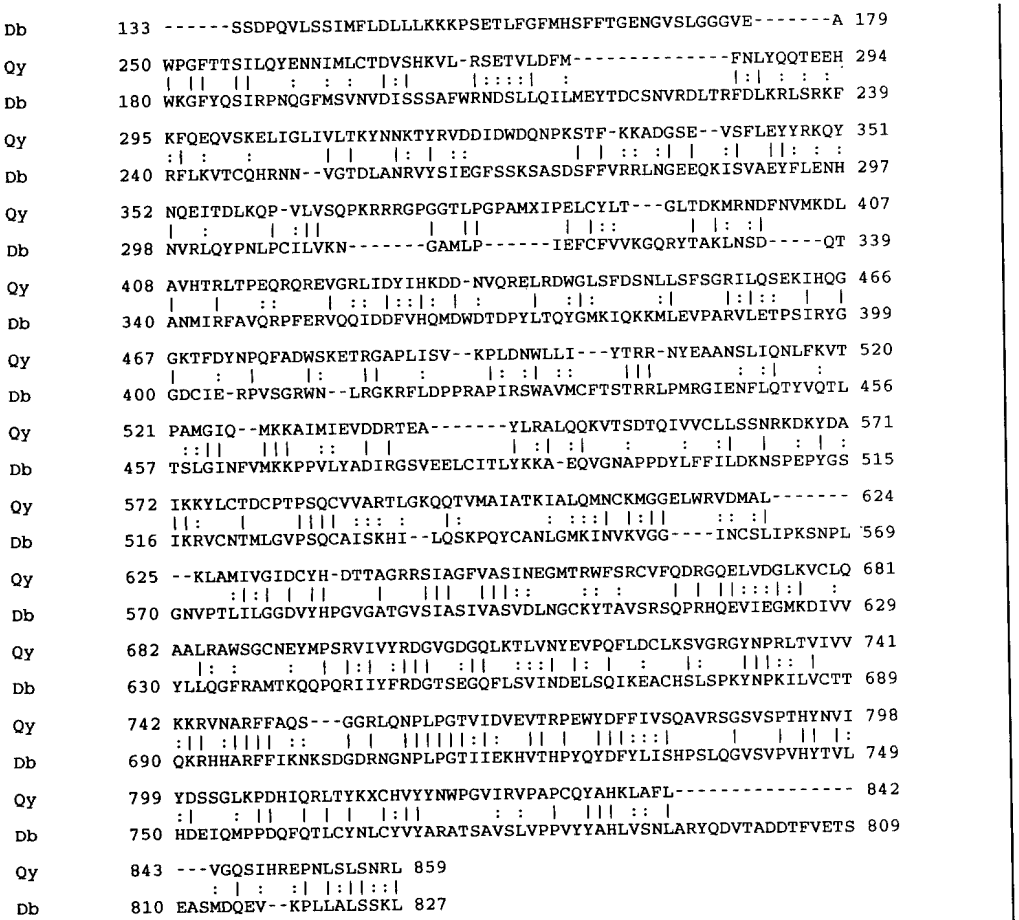
RESULT 3
T41568
argonaute-like protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C;Accession: T41568
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.
submitted to the EMBL Data Library, May 1998
A;Reference number: Z21991
A;Accession: T41568
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-834 <WOO>
A;Cross-references: EMBL:AL023705; PIDN:CAA19275.1; GSPDB:GN00068; SPDB:SPCC736.11
A;Experimental source: strain 972h-; cosmid c736
C;Genetics:
A;Gene: SPDB:SPCC736.11
A;Map position: 3
A;Introns: 43/3
C;Superfamily: rabbit translation initiation factor eIF-2C

 Query Match 12.0%; Score 541.5; DB 2; Length 834;
 Best Local Similarity 22.8%; Pred. No. 2.8e-32;
 Matches 196; Conservative 159; Mismatches 350; Indels 155; Gaps 33;

```
Qy        107 GSSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDL--------- 157
Db         16 GGLGKQITLKANFFQIISLPNETINQYHVIVGDGSRVPRKQSQLIWNSKEVKQYFGSSWM 75

Qy        158 ----IGRCHAF------DGTILF------LPKRLQHKVTE-------VFSQTRNGEHVRI 194
Db         76 NSVYDGRSMCWSKGDIADGTIKVNIGSESHPREIEFSIQKSSKINLHTLSQFVNSKY--- 132

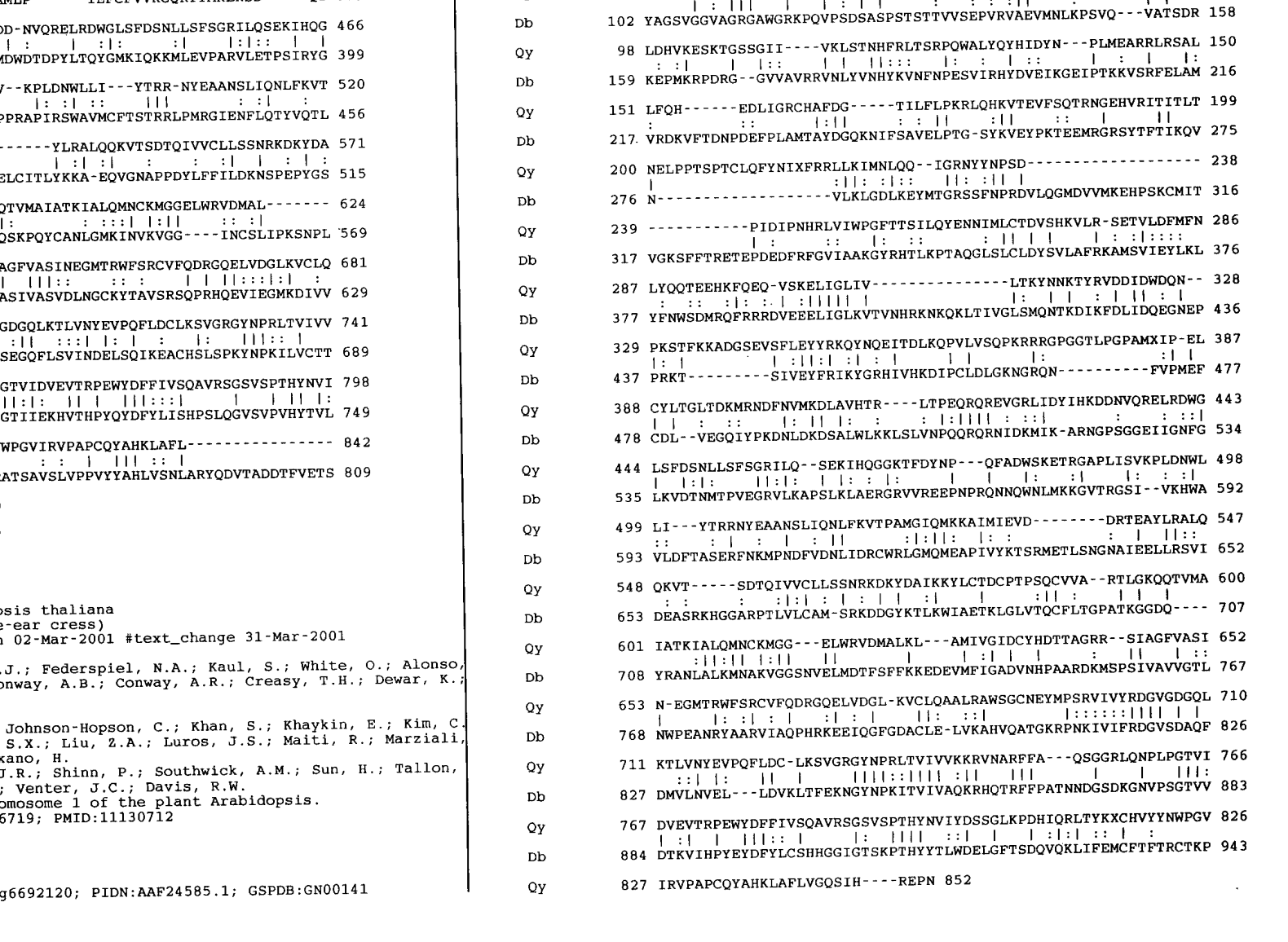
Qy        195 TITLTNELPPTSPTCLQFYNIXFRR-----LLKIMNLQQIGRNYYNPSDPIDIPNHRLVI 249
```



RESULT 4
H86438
protein T19E23.7 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: H86438
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H86438
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1014 <STO>
A;Cross-references: GB:AE005172; NID:g6692120; PIDN:AAF24585.1; GSPDB:GN00141

C;Genetics:
A;Gene: T19E23.7
A;Map position: 1

Query Match 11.5%; Score 522.5; DB 2; Length 1014;
Best Local Similarity 23.2%; Pred. No. 1e-30;
Matches 230; Conservative 164; Mismatches 397; Indels 199; Gaps 45;



```
Db        944 VSLVPPVYYADMVAFR-GRMYHEASSREKN 972
```

RESULT 5
T52134
Zwille protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Nov-2000
C;Accession: T52134
R;Laux, T.
submitted to the EMBL Data Library, January 1998
A;Reference number: Z25975
A;Accession: T52134
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-988 <LAU>
A;Cross-references: EMBL:AJ223508; PIDN:CAA11429.1
C;Genetics:
A;Gene: zwille
A;Map position: 5
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 10.9%; Score 495.5; DB 2; Length 988;
Best Local Similarity 23.5%; Pred. No. 1.1e-28;
Matches 223; Conservative 145; Mismatches 426; Indels 155; Gaps 36;

```
Qy          4 RARARARGRARGQETVQ-HVGAAASQQPGYIPPRPQQSPTEGDLVGRGRQRGMVVGATSK 62
Db         65 RSRRRNRG---GRKSDQGDVCMRPSSRPRK-PPPPSQT------------------TSS 101

Qy         63 SQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHFRL 122
Db        102 A-----------VSVATAGEIVAVNHQMQMGVRKNSNFAPRPGFGTLGTKCIVKANHF-L 149

Qy        123 TSRPQWALYQYHIDYNPLMEARRLRSAL------LFQHEDLIGRCHAFDGT--------- 167
Db        150 ADLPTKDLNQYDVTITPEVSSKSVNRAIIAELVRLYKESDLGRRLPAYDGRKSLYTAGEL 209

Qy        168 -----------------ILFLPKR-LQHKVTEVFSQTRNGEHVRITITLTNELPPTSPTC 209
Db        210 PFTWKEFSVKIVDEDDGIINGPKRERSYKVAIKFVARANMHH--LGEFLAGKRADCPQEA 267

Qy        210 LQFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIP---NHRLVIWPGFTTSILQYENNIML 266
Db        268 VQILDIVLRE-LSVKRFCPVGRSFFSPD--IKTPQRLGEGLESWCGFYQSIRPTQMGLSL 324

Qy        267 CTDV-SHKVLRSETVLDFMFNLYQQTEEHK-----FQEQVSKELIGLIVLTKYN---NKT 317
Db        325 NIDMASAAFIEPLPVIEFVAQLLGKDVLSKPLSDSDRVKIKKGLRGVKVEVTHRANVRRK 384

Qy        318 YRVDDIDWDQNPKSTFKKADGSEV-SFLEYYRKQYNQEITDLKQPVL-VSQPKRRRGPGG 375
Db        385 YRVAGLTTQPTRELMFPVDENCTMKSVIEYFQEMYGFTIQHTHLPCLQVGNQKK----AS 440

Qy        376 TLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNV 435
Db        441 YLPMEACKIVEGQRYT----KRLNEKQITALLKV-TCQRAEGQRNDILRTVQHNAYDQDP 495

Qy        436 QRELRDWGLSFDSNLLSFSGRILQSE--KIHQGGKTFDYNPQFADWSKETRGAPLISVKP 493
Db        496 Y--AKEFGMNISEKLASVEARILPAPWLKYHENGKEKDCLPQVGQWNMMNK--KMINGMT 551

Qy        494 LDNWLLIYTRRNYE--AANSLIQNLFKVTPAMGIQMKKAIMIEV----DDRTE-----AY 542
Db        552 VSRWACVNFSRSVQENVARGFCNELGQMCEVSGMEFNPEPVIPIYSARPDQVEKALKHVY 611

Qy        543 LRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVAR---TLGKQQTVM 599
Db        612 HTSMNKTKGKELELLLAILPDNNGSLYGDLKRICETELGLISQCCLTKHVFKISKQ---- 667

Qy        600 AIATKIALQMNCKMGGE--------LWRVDMALKLAMIV-GIDCYHDTTA--GRRSIAGF 648
```

```
Db        668 -YLADVSLKINVKMGGRNTVLVDAISCRIPLVSDIPTIIFGADVTHPENGEESSPSIAAV 726

Qy        649 VASIN-EGMTRWFSRCVFQDRGQELVDGLKVCLQAALR----------AWSGCNEYM 694
Db        727 VASQDWPEVTKYAGLVCAQAHRQELIQDLYKTWQDPVRGTVSGGMIRDLLISFRKATGQK 786

Qy        695 PSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFA--- 751
Db        787 PLRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEPNYQPPVTFIVVQKRHHTRLFANNH 846

Qy        752 ---QSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDH 808
Db        847 RDKNSTDRSGNILPGTVVDTKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDENNFTADG 906

Qy        809 IQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSN 857
Db        907 IQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRA--RFYLEPEIMQDN 953

RESULT 6
T32079
hypothetical protein T07D3.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C;Accession: T32079
R;Lamar, B.; Wamsley, P.; Twyman, B.
submitted to the EMBL Data Library, July 1997
A;Description: The sequence of C. elegans cosmid T07D3.
A;Reference number: Z21121
A;Accession: T32079
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-891 <LAM>
A;Cross-references: EMBL:AF016682; PIDN:AAB66187.1; GSPDB:GN00020; CESP:T07D3.7
A;Experimental source: strain Bristol N2; clone T07D3
C;Genetics:
A;Gene: CESP:T07D3.7
A;Map position: 2
A;Introns: 34/2; 92/2; 206/1; 560/1; 879/3
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 10.8%; Score 489.5; DB 2; Length 891;
Best Local Similarity 23.6%; Pred. No. 2.5e-28;
Matches 201; Conservative 141; Mismatches 332; Indels 177; Gaps 36;

```
Qy 107 GSSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDG 166
Db 52 GVEGRSILLRANHFAVRI-PGGSVQHYQIDVFPDKCPRRVNREVI----------G 96

Qy 167 TILFLPKRLQHKVTEVFSQTRN----------GEHVRITITLTNE----------LPPTS 206
Db 97 CLISSFSKYFTNIRPVYDGKRNMYTREPLPIGTEPMNFEVTLPGDSAVERKFSVTMKWIG 156

Qy 207 PTCL------------------QFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIPNHRL 247
Db 157 QVCLSALDDAMEGRVRQVPHEAVQSIDVILRH-LPSLKYTPVGRSFFTPPGVMK-PGMQM 214

Qy 248 ----------VIWPGFTTSILQYENNIMLCTDVSHKVL-RSETVLDFMFNLYQ------- 289
Db 215 HQESKLGGGREVWFGFHQSVRPSQWKMMLNIDVSATAFYRAMPVIEFVAEVLELPVQALA 274

Qy 290 ------QTEEHKFQEQVSKELIGL-IVLTKYN--NKTYRVDDIDWDQNPKSTF----KKA 336
Db 275 ERRALSDAQRVKF----TKEIRGLKIEITHCGAVRRKYRVCNVTRRPAQTQTFPLQLETG 330

Qy 337 DGSEVSFLEYYRKQYNQEITDLKQPVL----VSQPKRRRGPGGTLPGPAMXIPELC---- 388
Db 331 QTIECTVAKYFFDKYR---IQLKYPHLPCLQVGQEQKHT----YLP------PEVCDIVP 377

Qy 389 ---YLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQREL------ 439
Db 378 GQRCLKKLTDVQTSTM-----IKATARSAPE-REREICKLVS--------KAELSADPFA 423
```
```

```
Qy        440 RDWGLSFDSNLLSFSGRILQSEKIHQGG--KTFDYNPQFADWSKETRGAPLISVKPLDNW 497
Db        424 HEFGITINPAMTEVKGRVLSAPKLLYGGRHRATTALPNQGVW--DMRGKQFHTGMEVRTW 481

Qy        498 -LLIYTRRNYEAANSL---IQNLFKVT-----PAMGIQM--KKAIMIEVDDRTEAYLRAL 546
Db        482 AIACFAQQSHVKENDLRMFTTQLQRISTDAGMPIIGTPMFCKYASGVE---QVEPMFKYL 538

Qy        547 QQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIA 606
Db        539 KQTYSAIQLIVVVL--PGKTPIYAEVKRVGDTVLGIATQCVQAK--NAIRTTPQTLSNLC 594

Qy        607 LQMNCKMGGELWRVDMAL----------KLAMIVGIDCYHDTTAGRR--SIAGFVASINE 654
Db        595 LKMNVKLGG----VNSILLPNVRPRIFNEPVIFLGCDITHPAAGDTRKPSIAAVVGSMDA 650

Qy        655 GMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLV 714
Db        651 HPSRYAATVRVQQHRQEIITDLTYMVRELLVQFYRNTRFKPARIVVYRDGVSEGQLFNVL 710

Qy        715 NYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFF----AQSGGRLQNPLPGTVIDVEV 770
Db        711 QYELRAIREACVMLESGYQPGITFIAVQKRHHTRLFAADKADQVGKAFNIPPGTTVDVGI 770

Qy        771 TRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVP 830
Db        771 THPTEFDFFLCSHAGIQGTSRPSHYHVLWDDNDLTADELQQLTYQMCHTYVRCTRSVSIP 830

Qy        831 APCQYAHKLAF 841
Db        831 APAYYAHLVAF 841
```

RESULT 7
T22391
hypothetical protein F48F7.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T22391; T23234
R;Coles, L.
submitted to the EMBL Data Library, February 1996
A;Reference number: Z19559
A;Accession: T22391
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1000 <WIL>
A;Cross-references: EMBL:Z69661; PIDN:CAA93496.1; GSPDB:GN00028; CESP:F48F7.1
A;Experimental source: clone F48F7
R;Coles, L.
submitted to the EMBL Data Library, February 1996
A;Reference number: Z19712
A;Accession: T23234
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1000 <WI2>
A;Cross-references: EMBL:Z69663; PIDN:CAA93512.1; GSPDB:GN00028; CESP:F48F7.1
A;Experimental source: clone K02B9
C;Genetics:
A;Gene: CESP:F48F7.1
A;Map position: X
A;Introns: 70/3; 128/2; 185/2; 673/1
C;Superfamily: rabbit translation initiation factor eIF-2C

 Query Match 10.8%; Score 489.5; DB 2; Length 1000;
 Best Local Similarity 22.9%; Pred. No. 3e-28;
 Matches 201; Conservative 138; Mismatches 330; Indels 207; Gaps 36;

```
Qy        107 GSSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDG 166
Db        145 GVEGRSILLRANHFAVRI-PGGTIQHYQVDVTPDKCPRRVNREII----------S 189

Qy        167 TILFLPKRLQHKVTEVFSQTRNGEHVRITITLTNE-------LPPTSPTCLQF------- 212
```

```
Db        190 CLISAFSKYFTNIRPVYDGKRN-MYTREPLPIGRERMDFDVTLPGDSAVERQFSVSLKWV 248

Qy        213 ----------------YNIXFR---------RLLKIMNLQQIGRNYYNPSDPIDIPNHRL 247
Db        249 GQVSLSTLEDAMEGRVRQVPFEAVQAMDVILRHLPSLKYTPVGRSFFSP----PVPNASG 304

Qy        248 V--------------------------------IWPGFTTSILQYENNIMLCTDVSHK 273
Db        305 VMAGSCPPQASGAVAGGAHSAGQYHAESKLGGGREVWFGFHQSVRPSQWKMMLNIDVSAT 364

Qy        274 VL-RSETVLDFMFNLYQ------------QTEEHKFQEQVSKELIGL-IVLTKYN--NK 316
Db        365 AFYRSMPVIEFIAEVLELPVQALAERRALSDAQRVKF----TKEIRGLKIEITHCGQMRR 420

Qy        317 TYRVDDIDWDQNPKSTF----KKADGSEVSFLEYYRKQYNQEITDLKQPVL----VSQPK 368
Db        421 KYRVCNVTRRPAQTQTFPLQLETGQTIECTVAKYFYDKYR---IQLKYPHLPCLQVGQEQ 477

Qy        369 RRRGPGGTLPGPAMXIPELCYLT-------GLTDKMRNDFNVMKDLAVHTRLTPEQRQRE 421
Db        478 KHT----YLP------PEVCNIVPGQRCIKKLTDVQTSTM-----IKATARSAPE-RERE 521

Qy        422 VGRLI--------DYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKT-FDY 472
Db        522 ISNLVRKAEFSADPFAH----------EFGITINPAMTEVKGRVLSAPKLLYGGRTRATA 571

Qy        473 NPQFADWSKETRGAPLISVKPLDNW-LLIYTRRNYEAANSL---IQNLFKVTPAMGIQM- 527
Db        572 LPNQGVW--DMRGKQFHTGIDVRVWAIACFAQQQHVKENDLRMFTNQLQRISNDAGMPIV 629

Qy        528 ------KKAIMIEVDDRTEAYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCP 581
Db        630 GNPCFCKYAVGVE---QVEPMFKYLKQNY-SGIQLVVVVL-PGKTPVYAEVKRVGDTVLG 684

Qy        582 TPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMAL----------KLAMIVG 631
Db        685 IATQCVQAK--NAIRTTPQTLSNLCLKMNVKLGG----VNSILLPNVRPRIFNEPVIFFG 738

Qy        632 IDCYHDTTAGRR--SIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSG 689
Db        739 CDITHPPAGDSRKPSIAAVVGSMDAHPSRYAATVRVQQHRQEIISDLTYMVRELLVQFYR 798

Qy        690 CNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARF 749
Db        799 NTRFKPARIVVYRDGVSEGQFFNVLQYELRAIREACMMLERGYQPGITFIAVQKRHHTRL 858

Qy        750 FA----QSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLK 805
Db        859 FAVDKKDQVGKAYNIPPGTTVDVGITHPTEFDFYLCSHAGIQGTSRPSHYHVLWDDNNLT 918

Qy        806 PDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAF 841
Db        919 ADELQQLTYQMCHTYVRCTRSVSIPAPAYYAHLVAF 954
```

RESULT 8
A84668
Argonaute (AGO1)-like protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: A84668
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: A84668
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-930 <STO>
A;Cross-references: GB:AE002093; NID:g3885334; PIDN:AAC77862.1; GSPDB:GN00139
C;Genetics:

A;Gene: At2g27040
A;Map position: 2
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 10.2%; Score 462.5; DB 2; Length 930;
Best Local Similarity 20.9%; Pred. No. 2.9e-26;
Matches 183; Conservative 166; Mismatches 338; Indels 189; Gaps 34;

```
Qy        107 GSSGIIVKLSTNHFRL-TSRPQWALYQYHI-----DYNPLMEARRLRSALLFQ-----HE 155
              |: |  : | ||||::  :  |   : | :     |  | :| : :   :| :     |
Db         63 GTRGQKIPLLTNHFKVDVANLQGHFFHYSVALFYDDGRP-VEQKGVGRKILDKVHQTYHS 121

Qy        156 DLIGRCHAFDG-----TILFLP-KRLQHKV--------TEVFSQTRNG------------ 189
              || |:  |:||      |   ||  ::   |          :: |    ||
Db        122 DLDGKEFAYDGEKTLFTYGALPSNKMDFSVVLEEVSATSKDFVSRANGNGSPNGNESPSD 181

Qy        190 ------------EHVRITITLTNELPPTSPTCLQFYNIXFR-----------RLLKIMNL 226
                         :: |: |:   ::|      ||       |           |:| |:
Db        182 GDRKRLRRPNRSKNFRVEISYAAKIP------LQALANAMRGQESENSQEAIRVLDIILR 235

Qy        227 QQIGRN---------YYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDV-SHKVLR 276
              |   |          ::|     :     ::   || :|    :  : | || :  :::
Db        236 QHAARQGCLLVRQSFFHNDPTNCEPVGGNILGCRGFHSSFRTTQGGMSLNMDVTTTMIIK 295

Qy        277 SETVLDFMFNLYQQTEEHKFQ-EQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFK- 334
                |:||:       : :     :  : |  | |     : : :::  :       : ||:
Db        296 PGPVVDFLIANQNARDPYSIDWSKAKRTLKNLRVKVSPSGQEFKITGLSDKPCREQTFEL 355

Qy        335 ---------KADGSEVSFLEYYRK------QYNQEITDLKQPVLVSQPKRRRGPGGTLPG 379
                      : : :||:  :|:|       ||: ::  :     | :|||
Db        356 KKRNPNENGEFETTEVTVADYFRDTRHIDLQYSADLPCIN----VGKPKR---------- 401

Qy        380 PAMXIPELC-------YLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKD 432
              |    |||        |   ||   |:       |   :|  |::|   : : :    |
Db        402 PTYIPLELCALVPLQRYTKALTTFQRS------ALVEKSRQKPQERMTVLSKAL----KV 451

Qy        433 DNVQRE--LRDWGLSFDSNLLSFSGRILQSEKIHQ--GGKTFDYNPQFADWSKETRGAPL 488
              |   |  ||  |:|  ||     ||:| : |:    | :||  | ::   :||
Db        452 SNYDAEPLLRSCGISISSNFTQVEGRVLPAPKLKMGCGSETFPRNGRWNFNNKE-----F 506

Qy        489 ISVKPLDNWLLIYTRRNYEA---ANSLIQNLFKVTPAMGIQMKKAIMIEVDD-------- 537
              :    :  |:::    |: |       :: :| |:  : ||::      :  :
Db        507 VEPTKIQRWVVV----NFSARCNVRQVVDDLIKIGGSKGIEIASPFQVFEEGNQFRRAPP 562

Qy        538 --RTEAYLRALQQKVTSDTQIVVCLLSSNR-KDKYDAIKKYLCTDCPTPSQCVVARTLGK 594
                | |   : :| |:    | ::|:|   :  | |   ||   |:   :||:
Db        563 MIRVENMFKDIQSKLPGVPQFILCVLPDKKNSDLYGPWKKKNLTEFGIVTQCMAPTRQPN 622

Qy        595 QQTVMAIATKIALQMNCKMGGELWRVDMALKL------------AMIVGIDCYHDTTAG 641
              | :    | : |::| |:||    ::  | :              :|:|:|  |  : |
Db        623 DQYL----TNLLLKINAKLGG----LNSMLSVERTPAFTVISKVPTIILGMDVSHG-SPG 673

Qy        642 RRSIAGFVASINEG----MTRWFSRCVFQDRGQELVDGL---------KVCLQAALRAWS 688
              :  :    | ::      :::: :    |    |::: |           :  :  ::
Db        674 QSDVPSIAAVVSSREWPLISKYRASVRTQPSKAEMIESLVKKNGTEDDGIIKELLVDFYT 733

Qy        689 GCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNAR 748
              |:  |  :|::|||| : |   ::| |: | ::  | :   :||:  ::| :|  : :
Db        734 SSNKRKPEHIIIFRDGVSESQFNQVLNIELDQIIEACKLLDANWNPKFLLLVAQKNHHTK 793

Qy        749 FFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDH 808
              ||  :   :|  |||:|| ::  |:  ||:: : |   |:  ||||:|:||  |   |
Db        794 FFQPTSP--ENVPPGTIIDNKICHPKNNDFYLCAHAGMIGTTRPTHYHVLYDEIGFSADE 851

Qy        809 IQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVG 844
              :| | :   :||      | | ||  |||  |  :|
Db        852 LQELVHSLSYVYQRSTSAISVVAPICYAHLAAAQLG 887
```

RESULT 9
JC6569
translation initiation factor eIF-2C - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)

C;Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 21-Jul-2000
C;Accession: PC6505; JC6569
R;Zou, C.; Zhang, Z.; Wu, S.; Osterman, J.C.
Gene 211, 187-194, 1998
A;Title: Molecular cloning and characterization of a rabbit eIF2C protein.
A;Reference number: JC6569; MUID:98267198; PMID:9602122
A;Accession: PC6505
A;Molecule type: protein
A;Residues: 336-359;380-409;694-711 <ZOU1>
A;Experimental source: liver
A;Accession: JC6569
A;Molecule type: mRNA
A;Residues: 1-813 <ZOU2>
A;Cross-references: GB:AF005355; NID:g3253158; PIDN:AAC24323.1; PID:g3253159
C;Superfamily: rabbit translation initiation factor eIF-2C
C;Keywords: liver

Query Match 10.2%; Score 460; DB 2; Length 813;
Best Local Similarity 23.2%; Pred. No. 3.6e-26;
Matches 191; Conservative 144; Mismatches 321; Indels 166; Gaps 36;

```
Qy        126 PQWALYQYHIDYNPLMEARRLRSAL---LFQH--EDLIG-RCHAFDG-----TILFLPKR 174
              |:  :| | :|  |    ||:   :   : ||    : | |   |||     | : ||
Db          4 PKIDIYHYELDIKPEKCPRRVNREIVEHMVQHFKAQIFGDRKPVFDGRKNLYTAMPLPIG 63

Qy        175 LQHKVTEVFSQTRNGEH----VRITI------------TLTNELPPTSPTCLQFYNIXF 217
              :    ||  |  ||     :::|                |:  ||      :|  ::
Db         64 REKVELEV---TLPGEGKDRIFKVSIKWVSCVSLQALHDALSGRLPSVPFETIQALDVVM 120

Qy        218 RRLLKIMNLQQIGRNYYNPSDPIDIP-NHRLVIWPGFTTSILQYENNIMLCTDVSHKVL- 275
              |  |  |    :||:::  |:     |      :| ||  |:      :||  |||
Db        121 RH-LPSMRYTPVGRSFFTASEGCSNPLGGGREVWFGFHQSVRPSLWKMMLNIDVSATAFY 179

Qy        276 RSETVLDFMFNL--YQQTEEHK------FQEQVSKELIGLIVLTKY---NNKTYRVDDID 324
              ::: |::|:  :  ::  || :       : : :||: || |   :      : ||| ::
Db        180 KAQPVIEFVCEVLDFKSIEEQQKPLTDSQRVKFTKEIKGLKVEITHCGQMKRKYRVCNVT 239

Qy        325 WDQNPKSTFKKADGS----EVSFLEYYRKQYNQEITDLKQPVL-VSQPKRRRGPGGTLPG 379
                        ||     |    | :  :|:: ::   :      | | | | ::
Db        240 RRPASHQTFPLQQESGQTVECTVAQYFKDRHKLVLRYPHLPCLQVGQEQKH--------- 290

Qy        380 PAMXIP-ELCYLTG-------LTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHK 431
                :| |:| :          |||   |  : |   :     |  |: || |: :|:
Db        291 --TYLPLEVCNIVAGQRCIKKLTD---NQTSTM--IRATARSAPD-RQEEISKLMR--SA 340

Qy        432 DDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWSKETRGAPLISV 491
              |    :|::|:     :   :||:||   |: ||:     |  |
Db        341 SFNTDPYVREFGIMVKDEMTDVTGRVLQPPSILYGGR----NKAIA-------------- 382

Qy        492 KPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKK-AIMIEVDDR--TEAYLRALQQ 548
              |:      ::  || :                ||::|  ||       |  || :|::  :
Db        383 TPVQG---VWDMRNKQFHT------------GIEIKVWAIACFAPQRQCTEVHLKSFTE 426

Qy        549 ---KVTSDT--------------------------------QIVVCLLSSNRKDKYDA 571
                |:: |                                 |:|| :|   :   |
Db        427 QLRKISRDAGMPIQGQPCFCKYAQGADSVGPMFRHLKNTYAGLQLVVVIL-PGKTPVYAE 485

Qy        572 IKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGG------ELWRVDMALK 625
              :|:   |     :|||  : :  |:|      : : |::| |:||          |  : :
Db        486 VKRVGDTVLGMATQCVQMKNV--QRTTPQTLSNLCLKINVKLGGVNNILLPQGRPPVFQQ 543

Qy        626 LAMIVGIDCYHDTTA-GRR-SIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQAA 683
              : :| |  |     |:: ||| | |::    |: :    |  ||::  |   ::
Db        544 PVIFLGADVTHPPAGDGKKPSIAAVVGSMDAHPNRYCATVRVQQHRQEIIQDLAAMVREL 603

Qy        684 LRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKK 743
              |  :      : |:|:| ||||| :|| : ::::|:     :    : : | | :| |||:|
Db        604 LIQFYKSTRFKPTRIIFYRDGVSEGQFQQVLHHELLAIREACIKLEKDYQPGITFIVVQK 663

Qy        744 RVNARFFA----QSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIY 799
              | : | |      :  |:  |   || :| ::| |  :||:: | |   |:  |:||:|::
Db        664 RHHTRLFCTDKNERVGKSGNIPAGTTVDTKITHPTEFDFYLCSHAGIQGTSRPSHYHVLW 723
```

```
Qy        800 DSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAF 841
              | :    | :| |||: || |       : :|||  ||| :||
Db        724 DDNRFSSDELQILTYQLCHTYVRCTRSVSIPAPAYYAHLVAF 765


RESULT 10
A84678
Argonaute (AGO1)-like protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: A84678
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: A84678
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-997 <STO>
A;Cross-references: GB:AE002093; NID:g4510428; PIDN:AAD21514.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g27880
A;Map position: 2
C;Superfamily: rabbit translation initiation factor eIF-2C

  Query Match             10.2%;  Score 460;  DB 2;  Length 997;
  Best Local Similarity   22.3%;  Pred. No. 5e-26;
  Matches  219;  Conservative  154;  Mismatches  425;  Indels  184;  Gaps   37;

Qy          3 GRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQS--------------------PT 42
              | :| |  ||   |  |  |  |  ||      |||
Db         11 GASRGRGGGRRSDQRQDQSSGQVA--WPGL-----QQSYGGRGGSVSAGRGRGNVGRGEN 63

Qy         43 EGDL----------VGRGRQRGMV------VGATSKSQELQISAGFQELSLAERGGRRRD 86
               |||          |  || || :      | ::||:  :  |:  :|    |      :
Db         64 TGDLTATQVPVASAVSGGRGRGNIGDPTFSVASSSKTVSV-ASSSKEESKNTEVSETMSN 122

Qy         87 FHDXGVNTRQNLDHVKESKT----------GSSGIIVKLSTNHFRLTSRPQWALYQYHI 135
                    |:  :  :  : :           |: |  | :  ||| |       || | :
Db        123 LQITSTETKPEMTSLPPASSKAVTFPVRPGRGTLGKKVMVRANHF-LVQVADRDLYHYDV 181

Qy        136 DYNPLMEA----RRLRSALLFQHED--LIGRCHAFD--------GTILFLPKRLQHKVTE 181
              || : :    | :   |:  ::|  | |:  |:|        | : |  |     : |
Db        182 SINPEVISKTVNRNVMKLLVKNYKDSHLGGKSPAYDGRKSLYTAGPLPFDSKEFVVNLAE 241

Qy        182 VFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRR-------LLKIMNL-------- 226
                :   :|:     : : |    ||    |     |:         ::::::
Db        242 KRADGSSGKDRPFKVAVKN---VTSTDLYQLQQFLDRKQREAPYDTIQVLDVVLRDKPSN 298

Qy        227 --QQIGRNYYNPSDPIDIPNHR------LVIWPGFTTSILQYENNIMLCTDVSHKVLRSE 278
                 :||:::: |   |  : |      :  | |:  |:   :  : | |||  :
Db        299 DYVSVGRSFFHTSLGKDARDGRGELGDGIEYWRGYFQSLRLTQMGLSLNIDVSARSFYEP 358

Qy        279 TVL-DFM---FNLYQQTEEHKFQE--QVSKELIGL-IVLTKYN-NKTYRVDDIDWDQNPK 330
              |: ||:    |:       :  :  :| | |  | : |  :|  |: ::  |      :
Db        359 IVVTDFISKFLNIRDLNRPLRDSDRLKVKKVLRTLKVKLLHWNCTKSAKISGISSLPIRE 418

Qy        331 STFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMXIPELC-- 388
                |  | || : ::|: ::|| :      | : :          |:   |     |||
Db        419 LRFTLEDKSEKTVVQYFAEKYNYRVKYQALPAIQT---------GSDTRPVYLPMELCQI 469

Qy        389 -----YLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWG 443
                    |   | :|       :  |    |  |  |:  :  |:   : :|:: :|   :|
Db        470 DEGQRYTKRLNEKQ------VTALLKATCQRPPDRENSIKNLVVKNNYNDDLSKE---FG 520

Qy        444 LSFDSNLLSFSGRILQSE--KIHQGGKTFDYNPQFADWS----KETRGAPLISVKPLDNW 497
              :|  : | |   |:|     | |  ||    ||:   |:    |   || : |      |
Db        521 MSVTTQLASIEARVLPPPMLKYHDSGKEKMVNPRLGQWNMIDKKMVNGAKVTS------W 574

Qy        498 LLI--YTRRNYEAANSLIQNLFKVTPAMGIQMKKA-----IMIEVDDRTEAYLRALQQKV 550
                :   || :       : |  :  : |:: |       |    :   || |     |

Db        575 TCVSFSTRIDRGLPQEFCKQLIGMCVSKGMEFKPQPAIPFISCPPEHIEEALLDI--HKR 632

Qy        551 TSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMN 610
                |::: :|  :    |  ||:   |:    |||   | : |          :||::|
Db        633 APGLQLLIVIL-PDVTGSYGKIKRICETELGIVSQCCQPRQVNKLN--KQYMENVALKIN 689

Qy        611 CKMGGELWRVDMALKL---------AMIVGIDCYHDTTA--GRRSIAGFVASIN-EGMTR 658
              | ||    :: |::           :|:| |  |          |||  |||::   : :
Db        690 VKTGGRNTVLNDAIRRNIPLITDRPTIIMGADVTHPQPGEDSSPSIAAVVASMDWPEINK 749

Qy        659 WFSRCVFQDRGQELVDGLKVCLQAALR------------AWSGCNEYMPSRVIVYRDGV 705
              :      |   :|::  |   :|   |             |:      :| |:| |||||
Db        750 YRGLVSAQAHREEIIQDLYKLVQDPQRGLVHSGLIREHFIAFRRATGQIPQRIIFYRDGV 809

Qy        706 GDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGG------RLQN 759
              :||   :: :|:       |:   | ||:| ::|:|| : | | :  |      :  |
Db        810 SEGQFSQVLLHEMTAIRKACNSLQENYVPRVTFVIVQKRHHTRLFPEQHGNRDMTDKSGN 869

Qy        760 PLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHV 819
              ||||:| ::  |  :||:: | |   |:  | ||:|: | :|   | :| ||   |:
Db        870 IQPGTVVDTKICHPNEFDFYLNSHAGIQGTSRPAHYHVLLDENGFTADQLQMLTNNLCYT 929

Qy        820 YYNWPGVIRVPAPCQYAHKLAF 841
              |      : :  |  |||  ||
Db        930 YARCTKSVSIVPPAYYAHLAAF 951


RESULT 11
T23164
hypothetical protein T22B3.2a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T23164; T25099
R;Cottage, A.
submitted to the EMBL Data Library, January 1996
A;Reference number: Z19701
A;Accession: T23164
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1032 <WIL>
A;Cross-references: EMBL:Z68750; PIDN:CAA92969.1; GSPDB:GN00022; CESP:T22B3.2a
A;Experimental source: clone K01A6
R;Lennard, N.
submitted to the EMBL Data Library, December 1995
A;Reference number: Z19981
A;Accession: T25099
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1032 <WI2>
A;Cross-references: EMBL:Z68300; PIDN:CAA92618.1; GSPDB:GN00022; CESP:T22B3.2a
A;Experimental source: clone T22B3
C;Genetics:
A;Gene: CESP:T22B3.2a
A;Map position: 4
A;Introns: 19/1; 71/3; 129/2; 424/2; 496/3; 607/1; 666/1; 753/3; 795/2; 961/2; 1003/1
C;Superfamily: rabbit translation initiation factor eIF-2C

  Query Match             9.5%;  Score 428.5;  DB 2;  Length 1032;
  Best Local Similarity   21.6%;  Pred. No. 1.2e-23;
  Matches  200;  Conservative  146;  Mismatches  357;  Indels  225;  Gaps   36;

Qy        107 GSSGIIVKLSTNHFRL-TSRPQWALYQYHID-YNPLMEARRL---RSALLF-----QHED 156
              |: |  : : :| | :    |:  : |||:: ::|    |:|      ::|     | :
Db         88 GTIGRKIPVKSNFFAVDLKNPKMVVVQYHVEVHHP--GCRKLDKDEMRIIFWKAVSDHPN 145

Qy        157 LIGR--CHAFDGT-ILFLPKRLQ--------HKVTEVFSQTRNGEHVRITI--------- 196
              :      |:||   |:   ||:                |     | : |  |
Db        146 IFHNKFALAYDGAHQLYTVARLEFPDDQGSVRLDCEATLPKDNRDRTRCAISIQNVGPVL 205

Qy        197 -----TLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYN----------PSDPI 240
                  | || |     | :|  :|  |: |    |:   |:|                :
Db        206 LEMQRTRTNNLDERVLTPIQILDIICRQSLTCPLLKN-SANFYTWKSSCYRIPTAAGQAL 264
```

```
Qy        241 DIPNHRLVIWPGFTTSILQYEN-NIMLCTDVSHKVLRSE--TVLDFMFN---------- 286
              |:   :  :| || :|     |   :|  ||:|       ||| || :
Db        265 DLEGGK-EMWTGFFSSAHIASNYRPLLNIDVAHTAFYKTRITVLQFMCDVLNERTSKPNR 323

Qy        287 ------------------------------------------------LYQQTEEH 294
                                                               | : |:
Db        324 NNPRGPGAPGGYRGGRGGARGGSYQNFGNRGPPGANVRDDFGGNGLTFTMDTLSRDTQLS 383

Qy        295 KFQEQVSKELI-GLIVLTKYNNKTYRVDDIDWDQNP--KSTFKKAD--GSEV--SFLEYY 347
               |: ::  : | |: :   :     ||  ::  | |  |  |:  |  | :|  |  :|:
Db        384 SFETRIFGDSIRGMKIRATHRPNAIRVYKVNSLQLPADKLMFQGIDEEGRQVVCSVADYF 443

Qy        348 RKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMXIP-ELC-------YLTGLTDKMRN 399
              ::|      || | |   |    ||   |  : :| | |       |   :|:|  :
Db        444 SEKYG----PLKYPKL---PCLHVGP----PTRNIFLPMEHCLIDSPQKYNKKMTEKQTS 492

Qy        400 DFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQR-----------ELRDWGLSFDS 448
              ::|  ||                |   ::| :::             |:::|::  |
Db        493 --AIIKAAAV----------------DATQREDRIKQLAAQASFGTDPFLKEFGVAVSS 533

Qy        449 NLLSFSGRILQSEKIHQGGKTFDYN----PQFADWSKETRGAPLISVKPLDNWLLIYTRR 504
              ::  | |::|   |  ||      |    |:   || : :   : :      : : :  |
Db        534 QMIETSARVIQPPPIMFGGNNRSINPVVFPKDGSWSMDHQTLYMPATCRSYSMIALVDPR 593

Qy        505 NYEAANSLIQNLFKVTPAMGIQMKK--------------AIMIEVDDRTEAYLRALQQK 549
              :  :  :  |:|     |||:   :                :  |: |        : :
Db        594 DQTSLQTFCQSLTMKATAMGMNFPRWPDLVKYGRSKEDVCTLFTEIAD---------EYR 644

Qy        550 VTSDT-QIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQ 608
              ||:     :: :| |   | |  :|:          ||||: : : :     |    | |:
Db        645 VTNTVCDCIIVVLQSKNSDIYMTVKEQSDIVHGIMSQCVLMKNVSR--PTPATCANIILK 702

Qy        609 MNCKMGG-----------ELWRVDMALKLAMIVGIDCYHDTTAGRR----SIAGFVASIN 653
              :| ||||           : ||      |:||||  | | |  |    |:|  ||:::
Db        703 LNMKMGGINSRIVADQITNKYLVDQP---TMVVGIDVTHPTQAEMRMNMPSVAAIVANVD 759

Qy        654 EGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTL 713
                :  :    | : :| |  |   ::  :  :    :  |:|:|||||| :||   :
Db        760 LLPQSYGANVKVQKKCRESVVYLLDAIRERIITFYRHTKQKPARIIVYRDGVSEGQFSEV 819

Qy        714 VNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQ----SGGRLQNPLPGTVIDVE 769
              :  |:        ::   : | :| |||:|| :|| | :      |: :|  ||| :|
Db        820 LREEIQSIRTACLAIAEDFRPPITYIVVQKRHHARIFCKFPNDMVGKAKNVPPGTTVDTG 879

Qy        770 VTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRV 829
              :  || :||:: |     |:  |  |:|: |       | || :||  || |      : :
Db        880 IVSPEGFDFYLCSHYGVQGTSRPARYHVLLDECKFTADEIQNITYGMCHTYGRCTRSVSI 939

Qy        830 PAPCQYAHKLAFLVGQSIHREPNLSLSN 857
              | |  ||  :|      | |:  |: :|
Db        940 PTPVYYADLVATRARCHIKRKLGLADNN 967
```

RESULT 12
T23165
hypothetical protein T22B3.2b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T23165; T25100
R;Cottage, A.
submitted to the EMBL Data Library, January 1996
A;Reference number: Z19701
A;Accession: T23165
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1035 <WIL>
A;Cross-references: EMBL:Z68750; PIDN:CAA92970.1; GSPDB:GN00022; CESP:T22B3.2b
A;Experimental source: clone K01A6
R;Lennard, N.
submitted to the EMBL Data Library, December 1995
A;Reference number: Z19981
A;Accession: T25100
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1035 <WI2>
A;Cross-references: EMBL:Z68300; PIDN:CAA92619.1; GSPDB:GN00022; CESP:T22B3.2b
A;Experimental source: clone T22B3
C;Genetics:
A;Gene: CESP:T22B3.2b
A;Map position: 4
A;Introns: 19/1; 74/3; 132/2; 427/2; 499/3; 610/1; 669/1; 756/3; 798/2; 964/2; 1006/1
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 9.5%; Score 428.5; DB 2; Length 1035;
Best Local Similarity 21.6%; Pred. No. 1.3e-23;
Matches 200; Conservative 146; Mismatches 357; Indels 225; Gaps 36;

```
Qy        107 GSSGIIVKLSTNHFRL-TSRPQWALYQYHID-YNPLMEARRL---RSALLF-----QHED 156
              |: |  : : :| | :    |:  : |||:: ::|    |:|      ::|      | :
Db         91 GTIGRKIPVKSNFFAVDLKNPKMVVVQYHVEVHHP--GCRKLDKDEMRIIFWKAVSDHPN 148

Qy        157 LIGR--CHAFDGT-ILFLPKRLQ--------HKVTEVFSQTRNGEHVRITI--------- 196
              :       |:||   |:   ||:             |      | :  |  |
Db        149 IFHNKFALAYDGAHQLYTVARLEFPDDQGSVRLDCEATLPKDNRDRTRCAISIQNVGPVL 208

Qy        197 -----TLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYN----------PSDPI 240
                  | || |     | :|  :|  |: |    |:    |:|                :
Db        209 LEMQRTRTNNLDERVLTPIQILDIICRQSLTCPLLKN-SANFYTWKSSCYRIPTAAGQAL 267

Qy        241 DIPNHRLVIWPGFTTSILQYEN-NIMLCTDVSHKVLRSE--TVLDFMFN---------- 286
              |:   :  :| || :|     |   :|  ||:|       ||| || :
Db        268 DLEGGK-EMWTGFFSSAHIASNYRPLLNIDVAHTAFYKTRITVLQFMCDVLNERTSKPNR 326

Qy        287 ------------------------------------------------LYQQTEEH 294
                                                               | : |:
Db        327 NNPRGPGAPGGYRGGRGGARGGSYQNFGNRGPPGANVRDDFGGNGLTFTMDTLSRDTQLS 386

Qy        295 KFQEQVSKELI-GLIVLTKYNNKTYRVDDIDWDQNP--KSTFKKAD--GSEV--SFLEYY 347
               |: ::  : | |: :   :     ||  ::  | |  |  |:  |  | :|  |  :|:
Db        387 SFETRIFGDSIRGMKIRATHRPNAIRVYKVNSLQLPADKLMFQGIDEEGRQVVCSVADYF 446

Qy        348 RKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMXIP-ELC-------YLTGLTDKMRN 399
              ::|      || | |   |    ||   |  : :| | |       |   :|:|  :
Db        447 SEKYG----PLKYPKL---PCLHVGP----PTRNIFLPMEHCLIDSPQKYNKKMTEKQTS 495

Qy        400 DFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQR-----------ELRDWGLSFDS 448
              ::|  ||                |   ::| :::             |:::|::  |
Db        496 --AIIKAAAV----------------DATQREDRIKQLAAQASFGTDPFLKEFGVAVSS 536

Qy        449 NLLSFSGRILQSEKIHQGGKTFDYN----PQFADWSKETRGAPLISVKPLDNWLLIYTRR 504
              ::  | |::|   |  ||      |    |:   || : :   : :      : : :  |
Db        537 QMIETSARVIQPPPIMFGGNNRSINPVVFPKDGSWSMDHQTLYMPATCRSYSMIALVDPR 596

Qy        505 NYEAANSLIQNLFKVTPAMGIQMKK--------------AIMIEVDDRTEAYLRALQQK 549
              :  :  :  |:|     |||:   :                :  |: |        : :
Db        597 DQTSLQTFCQSLTMKATAMGMNFPRWPDLVKYGRSKEDVCTLFTEIAD---------EYR 647

Qy        550 VTSDT-QIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQ 608
              ||:     :: :| |   | |  :|:          ||||: : : :     |    | |:
Db        648 VTNTVCDCIIVVLQSKNSDIYMTVKEQSDIVHGIMSQCVLMKNVSR--PTPATCANIILK 705

Qy        609 MNCKMGG-----------ELWRVDMALKLAMIVGIDCYHDTTAGRR----SIAGFVASIN 653
              :| ||||           : ||      |:||||  | | |  |    |:|  ||:::
Db        706 LNMKMGGINSRIVADQITNKYLVDQP---TMVVGIDVTHPTQAEMRMNMPSVAAIVANVD 762

Qy        654 EGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTL 713
                :  :    | : :| |  |   ::  :  :    :  |:|:|||||| :||   :
Db        763 LLPQSYGANVKVQKKCRESVVYLLDAIRERIITFYRHTKQKPARIIVYRDGVSEGQFSEV 822

Qy        714 VNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQ----SGGRLQNPLPGTVIDVE 769
              :  |:        ::   : | :| |||:|| :|| | :      |: :|  ||| :|
Db        823 LREEIQSIRTACLAIAEDFRPPITYIVVQKRHHARIFCKFPNDMVGKAKNVPPGTTVDTG 882

Qy        770 VTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRV 829
              :  || :||:: |     |:  |  |:|: |       | || :||  || |      : :
Db        883 IVSPEGFDFYLCSHYGVQGTSRPARYHVLLDECKFTADEIQNITYGMCHTYGRCTRSVSI 942
```

```
Qy        830 PAPCQYAHKLAFLVGQSIHREPNLSLSN 857
              | |  ||  :|     | |:  |: :|
Db        943 PTPVYYADLVATRARCHIKRKLGLADNN 970
```

RESULT 13
S41013
hypothetical protein ZK757.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Feb-2000
C;Accession: S41013
R;Thomas, K.
submitted to the EMBL Data Library, December 1993
A;Reference number: S41011
A;Accession: S41013
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-958 <THO>
A;Cross-references: EMBL:Z29121
C;Genetics:
A;Introns: 52/2; 350/2; 422/3; 592/1; 679/3; 721/2; 887/2; 929/1
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 9.3%; Score 423; DB 2; Length 958;
Best Local Similarity 20.7%; Pred. No. 2.9e-23;
Matches 193; Conservative 153; Mismatches 357; Indels 228; Gaps 34;

```
Qy        107 GSSGIIVKLSTNHFRL-TSRPQWALYQYHID-YNPLMEARRL---RSALLF-----QHED 156
              |: |  : : :| | :    |:  : |||:: ::|    |:|      ::|       | :
Db         11 GTIGRQIPVKSNFFAMDLKNPKMVVIQYHVEIHHP--GCRKLDKDEMRIIFWKAVSDHPN 68

Qy        157 LIGR--CHAFDGT-ILFLPKRLQ--------HKVTEVFSQTRNGEHVRITI--------- 196
              :       |:||   |:   ||:          |      | :  |  |
Db         69 IFHNKFALAYDGAHQLYTVARLEFPDDQGSVRLDCEASLPKDNRDRTRCAISIQNVGPVL 128

Qy        197 -----TLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYN----------PSDPI 240
                  | || |     | :|  :|  |: |    |:    |:|                 :
Db        129 LEMQRTRTNNLDERVLTPIQILDIICRQSLTCPLLKN-SANFYTWKSSCYRIPTAAGQAL 187

Qy        241 DIPNHRLVIWPGFTTSILQYEN-NIMLCTDVSHKVLRSE--TVLDFMFN---------- 286
              |:   :  :| || :|     |   :|  ||:|        ||| || :
Db        188 DLEGGK-EMWTGFFSSAHIASNYRPLLNIDVAHTAFYKTRITVLQFMCDVLNERTSKPNR 246

Qy        287 ----------------------------------------------------LYQQT 291
                                                                   | : |
Db        247 NNPRGPGGPGGPGGYRGGRGGGRGGSYGNFGNRGPPGANVRDDFGGNGLTFTMDTLSRDT 306

Qy        292 EEHKFQEQVSKELI-GLIVLTKYNNKTYRVDDIDWDQNP--KSTFKKAD--GSEV--SFL 344
              :   |: ::  : | |: :    :      ||  ::  | |  |  |:  |  | :|  |
Db        307 QLSSFETRIFGDAIRGMKIRAAHRPNAIRVYKVNSLQLPADKLMFQGIDEEGRQVVCSVA 366

Qy        345 EYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVM 404
              :|: ::|      || | |   |    ||                          || |  |
Db        367 DYFSEKYG----PLKYPKL---PCLHVGP----------------------PTRNIFLPM 397

Qy        405 KDLAVHTRLTPEQRQREVGR----------LIDYIHKDDNVQR----------ELRDWG 443
              :  :    :|::  :::             :|   ::| :::           |:::|
Db        398 EHCLID---SPQKYNKKMSEKQTSAIIKAAAVDATQREDRIKQLAAQASFGTDPFLKEFG 454

Qy        444 LSFDSNLLSFSGRILQSEKIHQGGKTFDYN----PQFADWSKETRGAPLISVKPLDNWLL 499
              ::  | ::  : |::|   |  ||      |   |:  : : :   : :      : :
Db        455 VAVSSQMIQTTARVIQPPPIMFGGNNRSVNPVVFPKDGSWTMDNQTLYMPATCRSYSMIA 514

Qy        500 IYTRRNYEAANSLIQNLFKVTPAMGIQMKK--------------AIMIEVDDRTEAYLR 544
              :   |:  :  :  |:|     |||:   :                :  |: |
Db        515 LVDPRDQTSLQTFCQSLTMKATAMGMNFPRWPDLVKYGRSKEDVCTLFTEIAD------- 567

Qy        545 ALQQKVTSDT-QIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIAT 603
                : :||:     :: :| |   | |  :|:         ||||: : :  :   |
Db        568 --EYRVTNTVCDCIIVVLQSKNSDIYMTVKEQSDIVHGIMSQCVLMKNVSR--PTPATCA 623

Qy        604 KIALQMNCKMGGELWRVDMALKL---------AMIVGIDCYHDTTAGRR----SIAGFVA 650
```

```
                | |::| ||||   |: :| |:          |:||||  | | |  |    |:|  ||
Db        624 NIVLKLNMKMGGINSRI-VADKITNKYLVDQPTMVVGIDVTHPTQAEMRMNMPSVAAIVA 682

Qy        651 SINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQL 710
              :::     : :    | : :| |  |   ::  :  :    :  |:|:|||||||| :||
Db        683 NVDLLPQSYGANVKVQKKCRESVVYLLDAIRERIITFYRHTKQKPARIIVYRDGVSEGQF 742

Qy        711 KTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQ----SGGRLQNPLPGTVI 766
                ::  |:         ::   : | :| |||:|| :|| | :      |: :|  ||| :
Db        743 SEVLREEIQSIRTACLAIAEDFRPPITYIVVQKRHHARIFCKYQNDMVGKAKNVPPGTTV 802

Qy        767 DVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGV 826
              |  :  || :||:: |     |:  |  |:|: |       | || :||  || |
Db        803 DTGIVSPEGFDFYLCSHYGVQGTSRPARYHVLLDECKFTADEIQSITYGMCHTYGRCTRS 862

Qy        827 IRVPAPCQYAHKLAFLVGQSIHREPNLSLSN 857
              : :| |  ||  :|      : |:  |: :|
Db        863 VSIPTPVYYADLVATRARCHVKRKLGLADNN 893
```

RESULT 14
D88568
protein ZK757.3 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C;Accession: D88568
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A;Accession: D88568
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1040 <STO>
A;Cross-references: GB:chr_III; PIDN:CAA82941.1; PID:g3877004; GSPDB:GN00021; CESP:ZK
C;Genetics:
A;Gene: ZK757.3
A;Map position: 3
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 9.3%; Score 423; DB 2; Length 1040;
Best Local Similarity 20.7%; Pred. No. 3.3e-23;
Matches 193; Conservative 153; Mismatches 357; Indels 228; Gaps 34;

```
Qy        107 GSSGIIVKLSTNHFRL-TSRPQWALYQYHID-YNPLMEARRL---RSALLF-----QHED 156
              |: |  : : :| | :    |:  : |||:: ::|    |:|      ::|       | :
Db         93 GTIGRQIPVKSNFFAMDLKNPKMVVIQYHVEIHHP--GCRKLDKDEMRIIFWKAVSDHPN 150

Qy        157 LIGR--CHAFDGT-ILFLPKRLQ--------HKVTEVFSQTRNGEHVRITI--------- 196
              :       |:||   |:   ||:          |      | :  |  |
Db        151 IFHNKFALAYDGAHQLYTVARLEFPDDQGSVRLDCEASLPKDNRDRTRCAISIQNVGPVL 210

Qy        197 -----TLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYN----------PSDPI 240
                  | || |     | :|  :|  |: |    |:    |:|                 :
Db        211 LEMQRTRTNNLDERVLTPIQILDIICRQSLTCPLLKN-SANFYTWKSSCYRIPTAAGQAL 269

Qy        241 DIPNHRLVIWPGFTTSILQYEN-NIMLCTDVSHKVLRSE--TVLDFMFN---------- 286
              |:   :  :| || :|     |   :|  ||:|        ||| || :
Db        270 DLEGGK-EMWTGFFSSAHIASNYRPLLNIDVAHTAFYKTRITVLQFMCDVLNERTSKPNR 328

Qy        287 ----------------------------------------------------LYQQT 291
                                                                   | : |
Db        329 NNPRGPGGPGGPGGYRGGRGGGRGGSYGNFGNRGPPGANVRDDFGGNGLTFTMDTLSRDT 388

Qy        292 EEHKFQEQVSKELI-GLIVLTKYNNKTYRVDDIDWDQNP--KSTFKKAD--GSEV--SFL 344
              :   |: ::  : | |: :    :      ||  ::  | |  |  |:  |  | :|  |
Db        389 QLSSFETRIFGDAIRGMKIRAAHRPNAIRVYKVNSLQLPADKLMFQGIDEEGRQVVCSVA 448

Qy        345 EYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVM 404
              :|: ::|      || | |   |    ||                          || |  |
Db        449 DYFSEKYG----PLKYPKL---PCLHVGP----------------------PTRNIFLPM 479
```

```
Qy        405 KDLAVHTRLTPEQRQREVGR----------LIDYIHKDDNVQR----------ELRDWG 443
              :   :    :|::  :::             :|   ::| :::            |:::|
Db        480 EHCLID---SPQKYNKKMSEKQTSAIIKAAAVDATQREDRIKQLAAQASFGTDPFLKEFG 536

Qy        444 LSFDSNLLSFSGRILQSEKIHQGGKTFDYN----PQFADWSKETRGAPLISVKPLDNWLL 499
              ::  | ::  : |::|    |  ||      |    |:  |: : :  : :      : :
Db        537 VAVSSQMIQTTARVIQPPPIMFGGNNRSVNPVVFPKDGSWTMDNQTLYMPATCRSYSMIA 596

Qy        500 IYTRRNYEAANSLIQNLFKVTPAMGIQMKK---------------AIMIEVDDRTEAYLR 544
              :   |:  :  :  |:|     |||:   :              :  |: |
Db        597 LVDPRDQTSLQTFCQSLTMKATAMGMNFPRWPDLVKYGRSKEDVCTLFTEIAD------- 649

Qy        545 ALQQKVTSDT-QIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIAT 603
                : :||:     :: :| |   | |  :|:          ||||: : : :     |
Db        650 --EYRVTNTVCDCIIVVLQSKNSDIYMTVKEQSDIVHGIMSQCVLMKNVSR--PTPATCA 705

Qy        604 KIALQMNCKMGGELWRVDMALKL---------AMIVGIDCYHDTTAGRR----SIAGFVA 650
              | |:: | ||||   |: :| |:          |:|||  | | |  |    |:|  ||
Db        706 NIVLKLNMKMGGINSRI-VADKITNKYLVDQPTMVVGIDVTHPTQAEMRMNMPSVAAIVA 764

Qy        651 SINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQL 710
              :::     : :    | : :| |  |   ::  :  :     :  |:|:|||||||| :||
Db        765 NVDLLPQSYGANVKVQKKCRESVVYLLDAIRERIITFYRHTKQKPARIIVYRDGVSEGQF 824

Qy        711 KTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQ----SGGRLQNPLPGTVI 766
              ::  |:        ::   : | :| |||:|| :|| | :      |: :|  ||| :
Db        825 SEVLREEIQSIRTACLAIAEDFRPPITYIVVQKRHHARIFCKYQNDMVGKAKNVPPGTTV 884

Qy        767 DVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGV 826
              |  :  || :||:: |     |:  |  |:|: |      | || :||  || |
Db        885 DTGIVSPEGFDFYLCSHYGVQGTSRPARYHVLLDECKFTADEIQSITYGMCHTYGRCTRS 944

Qy        827 IRVPAPCQYAHKLAFLVGQSIHREPNLSLSN 857
              : :| |  ||  :|      : |:  |: :|
Db        945 VSIPTPVYYADLVATRARCHVKRKLGLADNN 975
```

RESULT 15
T01113
translation initiation factor eIF-2C homolog T21L14.12 - Arabidopsis thaliana
N;Alternate names: Argonaute (AGO1)-like protein
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 16-Feb-2001
C;Accession: T01113; D84739
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, December 1997
A;Description: Arabidopsis thaliana chromosome II BAC T21L14 genomic sequence.
A;Reference number: Z14209
A;Accession: T01113
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-887 <ROU>
A;Cross-references: EMBL:AC003033; NID:g2702261; PID:g2702284
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: D84739
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-887 <STO>
A;Cross-references: GB:AE002093; NID:g2702284; PIDN:AAB91987.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g32940; T21L14.12
A;Map position: 2
A;Introns: 61/3; 130/2; 203/2; 248/1; 281/3; 314/2; 367/3; 406/3; 442/3; 463/3; 501/3; 5
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 9.2%; Score 417.5; DB 2; Length 887;

Best Local Similarity 20.8%; Pred. No. 6.6e-23;
Matches 193; Conservative 163; Mismatches 353; Indels 219; Gaps 42;

```
Qy         65 ELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHFRLTS 124
              |   |     :|:       ||:    : ||: :        |::|  ::| |||| ::
Db          2 ETSSSLPLSPISIEPEQPSHRDY---DITTRRGV--------GTTGNPIELCTNHFNVSV 50

Qy        125 R-PQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVTEVF 183
              | |    ||| :             ::  :: |      | |||   : ::|  :: : :
Db         51 RQPDVVFYQYTV------------SITTENGD------AVDGT--GISRKLMDQLFKTY 89

Qy        184 SQTRNGEHVRI----TITLTNELP-------------------------PTSPTCL---- 210
              |   :|: :      |:     ||                        :| ||
Db         90 SSDLDGKRLAYDGEKTLYTVGPLPQNEFDFLVIVEGSFSKRDCGVSDGGSSSGTCKRSKR 149

Qy        211 ----QFYNIXFR----------------------------RLLKIMNLQQIGRN------ 232
                : | :                                |:| |:  ||
Db        150 SFLPRSYKVQIHYAAEIPLKTVLGTQRGAYTPDKSAQDALRVLDIVLRQQAAERGCLLVR 209

Qy        233 ---YYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVS-HKVLRSETVLDFMFNLY 288
                :::   |: :    : |  |  :|      : |  |||   :|    |::|:
Db        210 QAFFHSDGHPMKVGGGVIGI-RGLHSSFRPTHGGLSLNIDVSTTMILEPGPVIEFL-KAN 267

Qy        289 QQTEEHK---FQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKS----TFKKADGS-- 339
              |  |  :   : :  :| |  : |   : |  :::  |     | :    : |  ||
Db        268 QSVETPRQIDWIKVAAKMLKHMRVKATHRNMEFKI--IGLSSKPCNQQLFSMKIKDGERE 325

Qy        340 ----EVSFLEYYRKQYNQEITDLKQPVL-VSQPKRRRGPGGTLPGPAMXIPELCYLTGLT 394
                  |::  :|::: | : |:    | | | :| |          |     | | |  |
Db        326 VPIREITVYDYFKQTYTEPISSAYFPCLDVGKPDR----------PNYLPLEFCNLVSLQ 375

Qy        395 DKMRNDFNVMKDLAVHTR--LTPEQRQREVGR---LIDYIH-----KDDNVQRELRDWGL 444
                    | |:   |  |    ||: : |   | | :|     ||      |   |:
Db        376 -------RYTKPLSGRQRVLLVESSRQKPLERIKTLNDAMHTYCYDKDP----FLAGCGI 424

Qy        445 SFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRR 504
              | :  :   ||:|:   : : ||  |: |    |:   :   |:  : : :| ::
Db        425 SIEKEMTQVEGRVLKPPML-KFGKNEDFQPCNGRWNFNNK--MLLEPRAIKSWAIV---- 477

Qy        505 NYE---AANSLIQNLFKVTPAMGIQMKKAI-MIEVD---------DRTEAYLRALQQKVT 551
              |:      :: : : |       ||:: :   ::| |          :| |  :  :: |
Db        478 NFSFPCDSSHISRELISCGMRKGIEIDRPFALVEEDPQYKKAGPVERVEKMIATMKLKFP 537

Qy        552 SDTQIVVCLLSSNR-KDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMN 610
                ::|:|   :  | |   ||   |:    :||:    :  |       | : |::|
Db        538 DPPHFILCILPERKTSDIYGPWKKICLTEEGIHTQCICPIKISDQ-----YLTNVLLKIN 592

Qy        611 CKMGG--ELWRVDMALKL-------AMIVGIDCYHDTTAGRRSIAGFVASINEG----MT 657
               |:||   |  :: :  :          :|:|:|  |    ||  :   | :         ::
Db        593 SKLGGINSLLGIEYSYNIPLINKIPTLILGMDVSHG-PPGRADVPSVAAVVGSKCWPLIS 651

Qy        658 RWFSRCVFQDRGQELVDGLKVCLQAALRAWSG-CNEYM----------PSRVIVYRDGV 705
              |: :    |    |::| |   ::   :  :|  ||             | ::|::||||
Db        652 RYRAAVRTQSPRLEMIDSLFQPIENTEKGDNGIMNELFVEFYRTSRARKPKQIIIFRDGV 711

Qy        706 GDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLPGTV 765
              : | : ::  || | :   : :|    |: |||| :|  : : |   |   :|   |||
Db        712 SESQFEQVLKIEVDQIIKAYQRLGESDVPKFTVIVAQKNHHTKLFQAKGP--ENVPAGTV 769

Qy        766 IDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQR----LTYKXCHVYY 821
              :| ::  |  |||:: : | : |:  | ||:|: |  |  || :|     |:|| :  :
Db        770 VDTKIVHPTNYDFYMCAHAGKIGTSRPAHYHVLLDEIGFSPDDLQNLIHSLSYKLLNSIF 829

Qy        822 NWPGVIRV----PAPCQYAHKLAFLVGQ 845
              |   :: |     || :|||  |  | |
Db        830 NVSSLLCVFVLSVAPVRYAHLAAAQVAQ 857
```

Search completed: July 22, 2003, 07:05:53
Job time : 21.1325 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2003, 22:53:12 ; Search time 10.0779 Seconds
(without alignments)
3547.609 Million cell updates/sec

Title: US-09-873-737A-4
Perfect score: 4528
Sequence: 1 MTGRARARARGRARGQETVQ.....VGQSIHREPNLSLSNRLYYL 862

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	541.5	12.0	834	1	IF2C_SCHPO	O74957 schizosacch
2	506.5	11.2	988	1	PINH_ARATH	Q9xgw1 arabidopsis
3	499.5	11.0	1048	1	AGO1_ARATH	O04379 arabidopsis
4	496.5	11.0	857	1	I2C1_HUMAN	Q9ul18 homo sapien
5	481.5	10.6	863	1	I2C2_RAT	Q9qz81 rattus norv
6	460	10.2	813	1	I2C2_RABIT	O77503 oryctolagus
7	460	10.2	997	1	AGOL_ARATH	Q9sjk3 arabidopsis
8	423	9.3	1040	1	YO43_CAEEL	P34681 caenorhabdi
9	355	7.8	377	1	I2C2_HUMAN	Q9ukv8 homo sapien
10	237	5.2	1032	1	YQ53_CAEEL	Q09249 caenorhabdi
11	142.5	3.1	1384	1	RPOD_SINAL	Q9thv5 sinapis alb
12	122.5	2.7	618	1	NXF1_MOUSE	Q99jx7 mus musculu
13	122.5	2.7	1238	1	YN13_YEAST	P53840 saccharomyc
14	121.5	2.7	1376	1	RPOD_ARATH	P56764 arabidopsis
15	121	2.7	949	1	PMA6_ARATH	Q9sh76 arabidopsis
16	121	2.7	1388	1	RPOD_TOBAC	P38550 nicotiana t
17	118.5	2.6	1527	1	RPOD_MAIZE	P16025 zea mays (m
18	113.5	2.5	690	1	Y173_UREPA	Q9pqx0 ureaplasma
19	113.5	2.5	1547	1	RRPO_PMV	P20951 papaya mosa
20	113	2.5	1120	1	MFD_RICPR	O05955 rickettsia
21	112	2.5	944	1	DNL4_YEAST	Q08387 saccharomyc
22	111	2.5	1163	1	RPOD_PEA	P12227 pisum sativ
23	110.5	2.4	560	1	PHAC_PSEOL	P26496 pseudomonas
24	110.5	2.4	887	1	TOP1_BACAN	P40114 bacillus an
25	110	2.4	1343	1	RPOB_HAEIN	P43738 haemophilus
26	109.5	2.4	3144	1	VP13_YEAST	Q07878 saccharomyc
27	108.5	2.4	619	1	NXF1_HUMAN	Q9ubu9 homo sapien
28	107	2.4	4128	1	PRKD_MOUSE	P97313 mus musculu
29	107	2.4	6669	1	NEBU_HUMAN	P20929 homo sapien
30	106.5	2.4	102	1	GGC1_HUMAN	O60829 homo sapien
31	106.5	2.4	4568	1	DYHC_CAEEL	Q19020 caenorhabdi
32	105.5	2.3	1024	1	GCP5_HUMAN	Q96rt8 homo sapien
33	105.5	2.3	1597	1	RLR1_YEAST	P53552 saccharomyc
34	105	2.3	1805	1	HMW2_MYCGE	P47460 mycoplasma
35	105	2.3	2245	1	MYSJ_DICDI	P54697 dictyosteli
36	104	2.3	1066	1	HGPC_HAEIN	Q9x442 haemophilus
37	104	2.3	1168	1	DDX8_SCHPO	O42643 schizosacch
38	103.5	2.3	1086	1	SYI_RICPR	Q9zcu4 rickettsia
39	103.5	2.3	1886	1	POL_COYMV	P19199 commelina y
40	103	2.3	828	1	GYRA_STRPY	Q917q5 streptococc
41	103	2.3	1361	1	RPOD_SPIOL	P11704 spinacia ol
42	102.5	2.3	694	1	PPK_CAMCO	O32350 campylobact
43	102.5	2.3	1447	1	BUD4_YEAST	P47136 saccharomyc
44	102	2.3	1172	1	CU14_SCHPO	P41003 schizosacch
45	102	2.3	1616	1	VIT1_CAEEL	P55155 caenorhabdi

ALIGNMENTS

RESULT 1
IF2C_SCHPO
ID IF2C_SCHPO STANDARD; PRT; 834 AA.
AC O74957;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable eukaryotic translation initiation factor 2C (eIF2C) (eIF-2C).
GN SPCC736.11.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC initiation process (By similarity).
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL023705; CAA19275.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Initiation factor; Protein biosynthesis.
FT   DOMAIN      212    326       PAZ.
FT   DOMAIN      500    799       PIWI.
SQ   SEQUENCE   834 AA;  94438 MW;  F5CE397BF590D9E4 CRC64;

  Query Match           12.0%;  Score 541.5;  DB 1;  Length 834;
  Best Local Similarity  22.8%;  Pred. No. 3.4e-32;
  Matches  196;  Conservative  159;  Mismatches  350;  Indels  155;  Gaps   33;

Qy       107 GSSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDL--------- 157
Db        16 GGLGKQITLKANFFQIISLPNETINQYHVIVGDGSRVPRKQSQLIWNSKEVKQYFGSSWM 75

Qy       158 ----IGRCHAF------DGTILF------LPKRLQHKVTE-------VFSQTRNGEHVRI 194
Db        76 NSVYDGRSMCWSKGDIADGTIKVNIGSESHPREIEFSIQKSSKINLHTLSQFVNSKY--- 132

Qy       195 TITLTNELPPTSPTCLQFYNIXFRR-----LLKIMNLQQIGRNYYNPSDPIDIPNHRLVI 249
Db       133 ------SSDPQVLSSIMFLDLLLKKKPSETLFGFMHSFFTGENGVSLGGGVE-------A 179

Qy       250 WPGFTTSILQYENNIMLCTDVSHKVL-RSETVLDFM--------------FNLYQQTEEH 294
Db       180 WKGFYQSIRPNQGFMSVNVDISSSAFWRNDSLLQILMEYTDCSNVRDLTRFDLKRLSRKF 239

Qy       295 KFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTF-KKADGSE--VSFLEYYRKQY 351
Db       240 RFLKVTCQHRNN--VGTDLANRVYSIEGFSSKSASDSFFVRRLNGEEQKISVAEYFLENH 297

Qy       352 NQEITDLKQP-VLVSQPKRRRGPGGTLPGPAMXIPELCYLT---GLTDKMRNDFNVMKDL 407
Db       298 NVRLQYPNLPCILVKN-------GAMLP------IEFCFVVKGQRYTAKLNSD-----QT 339

Qy       408 AVHTRLTPEQRQREVGRLIDYIHKDD-NVQRELRDWGLSFDSNLLSFSGRILQSEKIHQG 466
Db       340 ANMIRFAVQRPFERVQQIDDFVHQMDWDTDPYLTQYGMKIQKKMLEVPARVLETPSIRYG 399

Qy       467 GKTFDYNPQFADWSKETRGAPLISV--KPLDNWLLI---YTRR-NYEAANSLIQNLFKVT 520
Db       400 GDCIE-RPVSGRWN--LRGKRFLDPPRAPIRSWAVMCFTSTRRLPMRGIENFLQTYVQTL 456

Qy       521 PAMGIQ--MKKAIMIEVDDRTEA-------YLRALQQKVTSDTQIVVCLLSSNRKDKYDA 571
Db       457 TSLGINFVMKKPPVLYADIRGSVEELCITLYKKA-EQVGNAPPDYLFFILDKNSPEPYGS 515

Qy       572 IKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMAL------- 624
Db       516 IKRVCNTMLGVPSQCAISKHI--LQSKPQYCANLGMKINVKVGG----INCSLIPKSNPL 569

Qy       625 --KLAMIVGIDCYH-DTTAGRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQ 681
Db       570 GNVPTLILGGDVYHPGVGATGVSIASIVASVDLNGCKYTAVSRSQPRHQEVIEGMKDIVV 629

Qy       682 AALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVV 741
Db       630 YLLQGFRAMTKQQPQRIIYFRDGTSEGQFLSVINDELSQIKEACHSLSPKYNPKILVCTT 689

Qy       742 KKRVNARFFAQS---GGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVI 798
Db       690 QKRHHARFFIKNKSDGDRNGNPLPGTIIEKHVTHPYQYDFYLISHPSLQGVSVPVHYTVL 749

Qy       799 YDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFL---------------- 842
Db       750 HDEIQMPPDQFQTLCYNLCYVYARATSAVSLVPPVYYAHLVSNLARYQDVTADDTFVETS 809

Qy       843 ---VGQSIHREPNLSLSNRL 859
Db       810 EASMDQEV--KPLLALSSKL 827


RESULT 2
PINH_ARATH
ID   PINH_ARATH     STANDARD;      PRT;   988 AA.
AC   Q9XGW1; O49256;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   PINHEAD protein (ZWILLE protein).
GN   AT5G43810 OR MQD19.17.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Landsberg erecta;
RX   MEDLINE=99094873; PubMed=9876176;
RA   Lynn K., Fernandez A., Aida M., Sedbrook J., Tasaka M., Masson P.,
RA   Barton M.K.;
RT   "The PINHEAD/ZWILLE gene acts pleiotropically in Arabidopsis
RT   development and has overlapping functions with the ARGONAUTE1 gene.";
RL   Development 126:469-481(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Landsberg erecta;
RX   MEDLINE=98169396; PubMed=9501101;
RA   Moussian B., Schoof H., Haecker A., Juergens G., Laux T.;
RT   "Role of the ZWILLE gene in the regulation of central shoot meristem
RT   cell fate during Arabidopsis embryogenesis.";
RL   EMBO J. 17:1799-1809(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: DEVELOPMENTAL PROTEIN REQUIRED FOR RELIABLE FORMATION OF
CC       PRIMARY AND AXILLARY SHOOT APICAL MERISTEMS. MAY BE A COMPONENT OF
CC       A HYPOTHETICAL MERISTEM FORMING COMPETENCE FACTOR.
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF154272; AAD40098.1; -.
DR   EMBL; AJ223508; CAA11429.1; -.
DR   EMBL; AB026651; BAB11310.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Developmental protein.
FT   DOMAIN      337    451       PAZ.
FT   DOMAIN      625    946       PIWI.
FT   CONFLICT    475    479       PRDRE -> AEGQR (IN REF. 2).
FT   CONFLICT    671    671       N -> D (IN REF. 2).
```

```
SQ   SEQUENCE   988 AA;  110867 MW;  32EBB349C613DA20 CRC64;

  Query Match             11.2%;  Score 506.5;  DB 1;  Length 988;
  Best Local Similarity   23.6%;  Pred. No. 1.7e-29;
  Matches  224;  Conservative  144;  Mismatches  426;  Indels  155;  Gaps   36;

Qy          4 RARARARGRARGQETVQ-HVGAAASQQPGYIPPRPQQSPTEGDLVGRGRQRGMVVGATSK 62
              |:| | ||   |::: |  |    | :|   || | |:                   ||
Db         65 RSRRRNRG---GRKSDQGDVCMRPSSRPRK-PPPPSQT------------------TSS 101

Qy         63 SQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHFRL 122
              :           :|:|  |       |   :  |:| :       |: |    :  ||| |
Db        102 A-----------VSVATAGEIVAVNHQMQMGVRKNSNFAPRPGFGTLGTKCIVKANHF-L 149

Qy        123 TSRPQWALYQYHIDYNPLMEARRLRSAL------LFQHEDLIGRCHAFDGT--------- 167
                 |   | || :   | : :: :  |:       |::  ||  |  |:||
Db        150 ADLPTKDLNQYDVTITPEVSSKSVNRAIIAELVRLYKESDLGRRLPAYDGRKSLYTAGEL 209

Qy        168 -----------------ILFLPKR-LQHKVTEVFSQTRNGEHVRITITLTNELPPTSPTC 209
                              |:  |||   :||   |   |  |  :   |  :
Db        210 PFTWKEFSVKIVDEDDGIINGPKRERSYKVAIKFVARANMHH--LGEFLAGKRADCPQEA 267

Qy        210 LQFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIP---NHRLVIWPGFTTSILQYENNIML 266
              :|  :|  | | :     :||::::|   |  |       |  | ||  ||   :  : |
Db        268 VQILDIVLRE-LSVKRFCPVGRSFFSPD--IKTPQRLGEGLESWCGFYQSIRPTQMGLSL 324

Qy        267 CTDV-SHKVLRSETVLDFMFNLYQQTEEHK-----FQEQVSKELIGLIVLTKYN---NKT 317
              |: |   :    |::|:  |  :    |       : :: | | |: |   :     :
Db        325 NIDMASAAFIEPLPVIEFVAQLLGKDVLSKPLSDSDRVKIKKGLRGVKVEVTHRANVRRK 384

Qy        318 YRVDDIDWDQNPKSTFKKADGSEV-SFLEYYRKQYNQEITDLKQPVL-VSQPKRRRGPGG 375
              |||  :      :  |   :   : | :||::: |   |      | | |   |:
Db        385 YRVAGLTTQPTRELMFPVDENCTMKSVIEYFQEMYGFTIQHTHLPCLQVGNQKK----AS 440

Qy        376 TLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNV 435
              ||  |  | |    |   |  |:  :   | | |   |  |: :: | : :   | :
Db        441 YLPMEACKIVEGQRYT----KRLNEKQITALLKV-TCQRPRDRENDILRTVQHNAYDQDP 495

Qy        436 QRELRDWGLSFDSNLLSFSGRILQSE--KIHQGGKTFDYNPQFADWSKETRGAPLISVKP 493
                 :::|::    | |   ||| :   | |: ||  |  ||   |:   :   :|:
Db        496 Y--AKEFGMNISEKLASVEARILPAPWLKYHENGKEKDCLPQVGQWNMMNK--KMINGMT 551

Qy        494 LDNWLLIYTRRNYE--AANSLIQNLFKVTPAMGIQMKKAIMIEV----DDRTE-----AY 542
              :  |  :   |: :   |       | ::    |::     :| :     |: |       |
Db        552 VSRWACVNFSRSVQENVARGFCNELGQMCEVSGMEFNPEPVIPIYSARPDQVEKALKHVY 611

Qy        543 LRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVAR---TLGKQQTVM 599
              :: :    : :::: :|  |    |  :|:   |:    ||| : :    : ||
Db        612 HTSMNKTKGKELELLLAILPDNNGSLYGDLKRICETELGLISQCCLTKHVFKISKQ---- 667

Qy        600 AIATKIALQMNCKMGGE--------LWRVDMALKLAMIV-GIDCYHDTTA--GRRSIAGF 648
                  ::|::| ||||             |: :   :  |: | |  |          |||
Db        668 -YLANVSLKINVKMGGRNTVLVDAISCRIPLVSDIPTIIFGADVTHPENGEESSPSIAAV 726

Qy        649 VASIN-EGMTRWFSRCVFQDRGQELVDGLKVCLQAALR------------AWSGCNEYM 694
              ||| :   :|::      |   |||:  |    |  :|             ::
Db        727 VASQDWPEVTKYAGLVCAQAHRQELIQDLYKTWQDPVRGTVSGGMIRDLLISFRKATGQK 786

Qy        695 PSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFA--- 751
              | |:| |||||  :||   :: ||:        |:   | | :| |||:|| : | ||
Db        787 PLRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEPNYQPPVTFIVVQKRHHTRLFANNH 846

Qy        752 ---QSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDH 808
                |  |  | |||||:| ::  |  :||:: | |    |:  | ||:|::| :    |
Db        847 RDKNSTDRSGNILPGTVVDTKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDENNFTADG 906

Qy        809 IQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSN 857
              || ||   |: |       : :  |  |||  ||      : || :   |
Db        907 IQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRA--RFYLEPEIMQDN 953


RESULT 3
AGO1_ARATH
```

```
ID   AGO1_ARATH     STANDARD;      PRT;  1048 AA.
AC   O04379;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Argonaute protein.
GN   AGO1 OR AT1G48410 OR F11A17.3.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia; TISSUE=Leaf;
RX   MEDLINE=98090460; PubMed=9427751;
RA   Bohmert K., Camus I., Bellini C., Bouchez D., Caboche M., Benning C.;
RT   "AGO1 defines a novel locus of Arabidopsis controlling leaf
RT   development.";
RL   EMBO J. 17:170-180(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).
CC   -!- FUNCTION: ESSENTIAL FOR PROPER DEVELOPMENT OF LEAVES AND FLORAL
CC       ORGANS, AND FORMATION OF AXILLARY MERISTEMS.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED AT LOW LEVELS.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT ALL DEVELOPMENTAL
CC       STAGES.
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U91995; AAC18440.1; -.
DR   EMBL; AC007932; AAD49755.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Developmental protein.
FT   DOMAIN      391    501       PAZ.
FT   DOMAIN      676    997       PIWI.
FT   DOMAIN       13    104       GLY-RICH.
```

```
SQ   SEQUENCE   1048 AA;  116190 MW;  3E5146343A09C541 CRC64;

  Query Match             11.0%;  Score 499.5;  DB 1;  Length 1048;
  Best Local Similarity   23.5%;  Pred. No. 6.2e-29;
  Matches  240;  Conservative  156;  Mismatches  425;  Indels  199;  Gaps   42;

Qy          3 GRARARARGRARG---QETVQHVGAAA-SQQP--------GY-IPPRPQQS--------- 40
              |      ||  ||    |   || |    : ||         ||  ||: ||
Db         23 GPVSGGGRGSQRGGFQQGGGQHQGGRGYTPQPQQGGRGGRGYGQPPQQQQQYGGPQEYQG 82

Qy         41 ------PTEGDLVGRGRQRG-------------------------------MVVGATS 61
                  | :|   | |  ||                                   |  |
Db         83 RGRGGPPHQGGRGGYGGGRGGGPSSGPPQRQSVPELHQATSPTYQAVSSQPTLSEVSPTQ 142

Qy         62 KSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHFR 121
                :   ::  |::||: |:|   :        :::      ::  | | ||    :  |||
Db        143 VPEPTVLAQQFEQLSV-EQGAPSQAIQPIPSSSKAFKFPMRPGK-GQSGKRCIVKANHF- 199

Qy        122 LTSRPQWALYQYHIDYNPLMEARRLRSALLFQHED------LIGRCHAFDG-----TILF 170
                |   |: | :   | : :| :  |:: |  |       |  |  |:||     |
Db        200 FAELPDKDLHHYDVTITPEVTSRGVNRAVMKQLVDNYRDSHLGSRLPAYDGRKSLYTAGP 259

Qy        171 LP------------------KRLQHKVTEVFSQTRNGEHVRITITLTNELPPTSPTCLQ 211
              ||                  :| : :   |       :   : : |  :        ||
Db        260 LPFNSKEFRINLLDEEVGAGGQRREREFKVVIKLVARADLHHLGMFLEGKQSDAPQEALQ 319

Qy        212 FYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIPNHR-----LVIWPGFTTSILQYENNIML 266
                :|  |  |       :||::|:|    ||   :     |  | || ||   :  : |
Db        320 VLDIVLRE-LPTSRYIPVGRSFYSP----DIGKKQSLGDGLESWRGFYQSIRPTQMGLSL 374

Qy        267 CTDVSHKV-LRSETVLDFMFNLYQQTEEHK-----FQEQVSKELIGL-IVLTKYNN--KT 317
                |:|    : :  |: |: :|  :      :       : :: | |: : :|   |  :
Db        375 NIDMSSTAFIEANPVIQFVCDLLNRDISSRPLSDADRVKIKKALRGVKVEVTHRGNMRRK 434

Qy        318 YRVDDIDWDQNPKSTFKKAD-GSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGT 376
              ||:  :      : ||   :  :: | :||: : |   |   : | |      |
Db        435 YRISGLTAVATRELTFPVDERNTQKSVVEYFHETYGFRIQHTQLPCLQVGNSNR------ 488

Qy        377 LPGPAMXIPELCYLT--GLTDKMRNDFNVMKDLAVHTRLTPEQRQREV---GRLIDYIHK 431
                |   |:| :       |  |:  :   | | |   |  |::::    :| ||  |
Db        489 ---PNYLPMEVCKIVEGQRYSKRLNERQITALLKV-TCQRPIDREKDILQTVQLNDYA-K 543

Qy        432 DDNVQRELRDWGLSFDSNLLSFSGRILQSE--KIHQGGKTFDYNPQFADWSKETRGAPLI 489
              |:  |    ::|:   ::| |   |||     | |: |:     ||   |:   :  :|
Db        544 DNYAQ----EFGIKISTSLASVEARILPPPWLKYHESGREGTCLPQVGQWNMMNK--KMI 597

Qy        490 SVKPLDNWLLIYTRRNYE--AANSLIQNLFKVTPAMGIQMK-KAIMIEVDDRTEAYLRAL 546
              :   ::||: |   |  :   | :  | | ::    |:    : ::  |  | |   : |
Db        598 NGGTVNNWICINFSRQVQDNLARTFCQELAQMCYVSGMAFNPEPVLPPVSARPEQVEKVL 657

Qy        547 QQK---VTS------DTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGK-QQ 596
              : :    ||      :  ::: :|  |    |  :|:   |:    ||| : : : |  :
Db        658 KTRYHDATSKLSQGKEIDLLIVILPDNNGSLYGDLKRICETELGIVSQCCLTKHVFKMSK 717

Qy        597 TVMAIATKIALQMNCKMGGELWRVDMAL---------KLAMIVGIDCYH-----DTTAGR 642
                ||    :||::| |:||    :  ||         :  :| | |  |     |::
Db        718 QYMA---NVALKINVKVGGRNTVLVDALSRRIPLVSDRPTIIFGADVTHPHPGEDSSP-- 772

Qy        643 RSIAGFVASIN-EGMTRWFSRCVFQDRGQELVD----------------GLKVCLQAALR 685
               |||  ||| :   :|::    |   |||:                 |:   |  | |
Db        773 -SIAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFKEWKDPQKGVVTGGMIKELLIAFR 831

Qy        686 AWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRV 745
                :|   : | |:| |||| :||   :: ||:         |:  || | :| :||:||
Db        832 RSTG---HKPLRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEAGYQPPVTFVVVQKRH 888

Qy        746 NARFFAQ------SGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIY 799
              : | |||      |  |  | |||||:| ::  |  :||::  | |   |:  | ||:|::
Db        889 HTRLFAQNHNDRHSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLW 948

Qy        800 DSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRL 859
              | :   | :| ||   |: |      : :  |  |||  ||      : ||  | |  :
Db        949 DENNFTADGLQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRA--RFYMEPETSDSGSM 1006
```

```
RESULT 4
I2C1_HUMAN
ID   I2C1_HUMAN     STANDARD;      PRT;   857 AA.
AC   Q9UL18;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Eukaryotic translation initiation factor 2C 1 (eIF2C 1) (eIF-2C 1)
DE   (Putative RNA-binding protein Q99).
GN   EIF2C1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20005943; PubMed=10534406;
RA   Koesters R., Adams V., Betts D., Moos R., Schmid M., Siermann A.,
RA   Hassam S., Weitz S., Lichter P., Heitz P.U., von Knebel Doeberitz M.,
RA   Briner J.;
RT   "Human eukaryotic initiation factor EIF2C1 gene: cDNA sequence,
RT   genomic organization, localization to chromosomal bands 1p34-p35, and
RT   expression.";
RL   Genomics 61:210-218(1999).
CC   -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC       initiation process.
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF093097; AAF00068.1; -.
DR   Genew; HGNC:3262; EIF2C1.
DR   MIM; 606228; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Initiation factor; Protein biosynthesis.
FT   DOMAIN      226    346       PAZ.
FT   DOMAIN      515    816       PIWI.
SQ   SEQUENCE   857 AA;  97214 MW;  1DBB524AE7CBAF66 CRC64;

  Query Match             11.0%;  Score 496.5;  DB 1;  Length 857;
  Best Local Similarity   23.5%;  Pred. No. 7.7e-29;
  Matches  199;  Conservative  160;  Mismatches  347;  Indels  141;  Gaps   37;

Qy        107 GSSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSAL---LFQH--EDLIG-R 160
              |: |  :||  |:| :   |:  :| | :|  |    ||:   :  : ||    : | |
Db         30 GTVGKPIKLLANYFEV-DIPKIDVYHYEVDIKPDKCPRRVNREVVEYMVQHFKPQIFGDR 88

Qy        161 CHAFDG-----TILFLP---KRLQHKVT-------EVFSQTRNGEHVRITI--------- 196
                 :||      |:  ||   :|:  :||        :|         :::|
Db         89 KPVYDGKKNIYTVTALPIGNERVDFEVTIPGEGKDRIF---------KVSIKWLAIVSWR 139

Qy        197 ----TLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIP-NHRLVIWP 251
                   | :   |    :|  ::  | |   |    :||::::| :    |      :|
Db        140 MLHEALVSGQIPVPLESVQALDVAMRHLAS-MRYTPVGRSFFSPPEGYYHPLGGGREVWF 198

Qy        252 GFTTSILQYENNIMLCTDVSHKVL-RSETVLDFMF------NLYQQTEEHKFQEQV--SK 302
              ||  |:      :||  |||     ::: |::||     |: :| :     ::|  :|
Db        199 GFHQSVRPAMWKMMLNIDVSATAFYKAQPVIEFMCEVLDIRNIDEQPKPLTDSQRVRFTK 258
```

```
Qy        303 ELIGLIVLTKY---NNKTYRVDDIDWDQNPKSTF----KKADGSEVSFLEYYRKQYNQEI 355
Db        259 EIKGLKVEVTHCGQMKRKYRVCNVTRRPASHQTFPLQLESGQTVECTVAQYFKQKYNLQ- 317

Qy        356 TDLKQPVL----VSQPKRRRGPGGTLPGPAMXIP-ELCYLTG-------LTDKMRNDFNV 403
Db        318 --LKYPHLPCLQVGQEQKH-----------TYLPLEVCNIVAGQRCIKKLTD---NQTST 361

Qy        404 MKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKI 463
Db        362 M--IKATARSAPD-RQEEISRLMK--NASYNLDPYIQEFGIKVKDDMTEVTGRVLPAPIL 416

Qy        464 HQGGKTFDY-NPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQN----LFK 518
Db        417 QYGGRNRAIATPNQGVW--DMRGKQFYNGIEIKVWAIACFAPQKQCREEVLKNFTDQLRK 474

Qy        519 VTPAMGIQMKK----AIMIEVDDRTEAYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKK 574
Db        475 ISKDAGMPIQGQPCFCKYAQGADSVEPMFRHL-KNTYSGLQLIIVIL-PGKTPVYAEVKR 532

Qy        575 YLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGG------ELWRVDMALKLAM 628
Db        533 VGDTLLGMATQCVQVKNVVK--TSPQTLSNLCLKINVKLGGINNILVPHQRSAVFQQPVI 590

Qy        629 IVGIDCYHDTTA-GRR-SIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRA 686
Db        591 FLGADVTHPPAGDGKKPSITAVVGSMDAHPSRYCATVRVQRPRQEIIEDLSYMVRELLIQ 650

Qy        687 WSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVN 746
Db        651 FYKSTRFKPTRIIFYRDGVPEGQLPQILHYELLAIRDACIKLEKDYQPGITYIVVQKRHH 710

Qy        747 ARFFA----QSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSS 802
Db        711 TRLFCADKNERIGKSGNIPAGTTVDTNITHPFEFDFYLCSHAGIQGTSRPSHYYVLWDDN 770

Qy        803 GLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAF-----------------LVG 844
Db        771 RFTADELQILTYQLCHTYVRCTRSVSIPAPAYYARLVAFRARYHLVDKEHDSGEGSHISG 830

Qy        845 QSIHREP 851
Db        831 QSNGRDP 837
```

```
RESULT 5
I2C2_RAT
ID   I2C2_RAT       STANDARD;      PRT;   863 AA.
AC   Q9QZ81;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2)
DE   (Golgi ER protein 95 kDa) (GERp95).
GN   EIF2C2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Hepatoma;
RX   MEDLINE=99443791; PubMed=10512872;
RA   Cikaluk D.E., Tahbaz N., Hendricks L.C., DiMattia G.E., Hansen D.,
RA   Pilgrim D., Hobman T.C.;
RT   "GERp95, a membrane-associated protein that belongs to a family of
RT   proteins involved in stem cell differentiation.";
RL   Mol. Biol. Cell 10:3357-3372(1999).
CC   -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC       initiation process.
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF195534; AAF12800.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Initiation factor; Protein biosynthesis.
FT   DOMAIN      239    352       PAZ.
FT   DOMAIN      521    822       PIWI.
SQ   SEQUENCE   863 AA;  97659 MW;  9A1E0D6B84F7B622 CRC64;

  Query Match           10.6%;  Score 481.5;  DB 1;  Length 863;
  Best Local Similarity   23.5%;  Pred. No. 1e-27;
  Matches  193;  Conservative  149;  Mismatches  353;  Indels  125;  Gaps   34;
```

```
Qy        107 GSSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSAL---LFQH--EDLIG-R 160
Db         36 GTTGRTIKLQANFFEM-DIPKIDIYHYELDIKPEKCPRRVNREIVEHMVQHFKTQIFGDR 94

Qy        161 CHAFDG-----TILFLPKRLQHKVTEVFSQTRNGEH----VRITI-------------TL 198
Db         95 KPVFDGRKNLYTAMPLPIGRDKVELEV---TLPGEGKDRIFKVSIKWVSCVSLQALHDAL 151

Qy        199 TNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIP-NHRLVIWPGFTTSI 257
Db        152 SGRLPSVPFETIQALDVVMRH-LPSMRYTPVGRSFFTASEGCSNPLGGGREVWFGFHQSV 210

Qy        258 LQYENNIMLCTDVSHKVL-RSETVLDFMFNL--YQQTEEHK------FQEQVSKELIGLI 308
Db        211 RPSLWKMMLNIDVSATAFYKAQPVIEFVCEVLDFKSIEEQQKPLTDSQRVKFTKEIKGLK 270

Qy        309 VLTKY---NNKTYRVDDIDWDQNPKSTFKKADGS----EVSFLEYYRKQYNQEITDLKQP 361
Db        271 VEITHCGQMKRKYRVCNVTRRPASHQTFPLQQESGQTVECTVAQYFKDRHKLVLRYPHLP 330

Qy        362 VL-VSQPKRRRGPGGTLPGPAMXIP-ELCYLTG-------LTDKMRNDFNVMKDLAVHTR 412
Db        331 CLQVGQEQKH-----------TYLPLEVCNIVAGQRCIKKLTD---NQTSTM--IRATAR 374

Qy        413 LTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDY 472
Db        375 SAPD-RQEEISKLMR--SASFNTDPYVREFGIMVKDEMTDVTGRVLQPPSILYGGR---- 427

Qy        473 NPQFADWSKETRGAPLISVKPLDN-----------WLLIYTRRNYEAA----NSLIQNLF 517
Db        428 --------NKAIATPVQGVWDMRNKQFHTGIEIKVWAIACFAPQRQCTEVHLKSFTEQLR 479

Qy        518 KVTPAMGIQMKK----AIMIEVDDRTEAYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIK 573
Db        480 KISRDAGMPIQGQPCFCKYAQGADSVEPMFRHL-KNTYAGLQLVVVIL-PGKTPVYAEVK 537

Qy        574 KYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGG------ELWRVDMALKLA 627
Db        538 RVGDTVLGMATQCVQMKNV--QRTTPQTLSNLCLKINVKLGGVNNILLPQGRPPVFQQPV 595

Qy        628 MIVGIDCYHDTTA-GRR-SIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALR 685
Db        596 IFLGADVTHPPAGDGKKPSIAAVVGSMDAHPNRYCATVRVQQHRQEIIQDLAAMVRELLI 655

Qy        686 AWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRV 745
Db        656 QFYKSTRFKPTRIIFYRDGVSEGQFQQVLHHELLAIREACIKLEKEYQPGITFIVVQKRH 715
```

```
Qy        746 NARFFA----QSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDS 801
              : | |     :  |:  |   || :| ::| |  :||:: | |   |:  |:||:|:::|
Db        716 HTRLFCTDKNERVGKSGNIPAGTTVDTKITHPTEFDFYLCSHAGIQGTSRPSHYHVLWDD 775

Qy        802 SGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAF 841
              :    | :| |||: || |      : :|||  ||| :||
Db        776 NRFSSDELQILTYQLCHTYVRCTRSVSIPAPAYYAHLVAF 815


RESULT 6
I2C2_RABIT
ID   I2C2_RABIT     STANDARD;      PRT;   813 AA.
AC   O77503;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2).
GN   EIF2C2.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Liver;
RX   MEDLINE=98267198; PubMed=9602122;
RA   Zou C., Zhang Z., Wu S., Osterman J.C.;
RT   "Molecular cloning and characterization of a rabbit eIF2C protein.";
RL   Gene 211:187-194(1998).
CC   -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC       initiation process.
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF005355; AAC24323.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Initiation factor; Protein biosynthesis.
FT   DOMAIN      189    302       PAZ.
FT   DOMAIN      471    772       PIWI.
SQ   SEQUENCE   813 AA;  92281 MW;  13BA0DB70E10688F CRC64;

  Query Match             10.2%;  Score 460;  DB 1;  Length 813;
  Best Local Similarity   23.2%;  Pred. No. 3.6e-26;
  Matches  191;  Conservative  144;  Mismatches  321;  Indels  166;  Gaps   36;

Qy        126 PQWALYQYHIDYNPLMEARRLRSAL---LFQH--EDLIG-RCHAFDG-----TILFLPKR 174
              |:  :| | :|  |    ||:   :   : ||    : | |   |||     | : || .
Db          4 PKIDIYHYELDIKPEKCPRRVNREIVEHMVQHFKAQIFGDRKPVFDGRKNLYTAMPLPIG 63

Qy        175 LQHKVTEVFSQTRNGEH----VRITI------------TLTNELPPTSPTCLQFYNIXF 217
              :   ||   |  ||      :::|              |:  ||      :|  ::
Db         64 REKVELEV---TLPGEGKDRIFKVSIKWVSCVSLQALHDALSGRLPSVPFETIQALDVVM 120

Qy        218 RRLLKIMNLQQIGRNYYNPSDPIDIP-NHRLVIWPGFTTSILQYENNIMLCTDVSHKVL- 275
              |  |  |    :||:::  |:    |      :| ||  |:      :||  |||
Db        121 RH-LPSMRYTPVGRSFFTASEGCSNPLGGGREVWFGFHQSVRPSLWKMMLNIDVSATAFY 179

Qy        276 RSETVLDFMFNL--YQQTEEHK------FQEQVSKELIGLIVLTKY---NNKTYRVDDID 324
              ::: |::|:  :  ::  || :       : : :||: || |   :     : ||| ::
Db        180 KAQPVIEFVCEVLDFKSIEEQQKPLTDSQRVKFTKEIKGLKVEITHCGQMKRKYRVCNVT 239

Qy        325 WDQNPKSTFKKADGS----EVSFLEYYRKQYNQEITDLKQPVL-VSQPKRRRGPGGTLPG 379
                    ||      |    | :  :|:: ::   :      | | | | ::
Db        240 RRPASHQTFPLQQESGQTVECTVAQYFKDRHKLVLRYPHLPCLQVGQEQKH--------- 290

Qy        380 PAMXIP-ELCYLTG-------LTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHK 431
              :| |:| :          |||   |  : |  :    |  |: || |: :|:
Db        291 --TYLPLEVCNIVAGQRCIKKLTD---NQTSTM--IRATARSAPD-RQEEISKLMR--SA 340

Qy        432 DDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWSKETRGAPLISV 491
              |    :|::|:     :   :||:||   |  ||:    |  |
Db        341 SFNTDPYVREFGIMVKDEMTDVTGRVLQPPSILYGGR----NKAIA------------- 382

Qy        492 KPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKK-AIMIEVDDR--TEAYLRALQQ 548
              |:     ::  || :               ||::|  ||  .   |  || :|::  :
Db        383 TPVQG---VWDMRNKQFHT------------GIEIKVWAIACFAPQRQCTEVHLKSFTE 426

Qy        549 ---KVTSDT---------------------------------QIVVCLLSSNRKDKYDA 571
                |:: |                                   |:|| :|   :   |
Db        427 QLRKISRDAGMPIQGQPCFCKYAQGADSVGPMFRHLKNTYAGLQLVVVIL-PGKTPVYAE 485

Qy        572 IKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGG------ELWRVDMALK 625
              :|:   |     :|||  : :  |:|     : :  |::| |:||         |  :  :
Db        486 VKRVGDTVLGMATQCVQMKNV--QRTTPQTLSNLCLKINVKLGGVNNILLPQGRPPVFQQ 543

Qy        626 LAMIVGIDCYHDTTA-GRR-SIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQAA 683
              : :| |  |      |:: |||  | |::    |: :     |   ||::  |   ::
Db        544 PVIFLGADVTHPPAGDGKKPSIAAVVGSMDAHPNRYCATVRVQQHRQEIIQDLAAMVREL 603

Qy        684 LRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKK 743
              |  :     : |:|:| ||||| :|| : ::::|:    :    : : | | :| |||:|
Db        604 LIQFYKSTRFKPTRIIFYRDGVSEGQFQQVLHHELLAIREACIKLEKDYQPGITFIVVQK 663

Qy        744 RVNARFFA----QSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIY 799
              | : | |     :  |:  |   || :| ::| |  :||:: | |   |:  |:||:|:::
Db        664 RHHTRLFCTDKNERVGKSGNIPAGTTVDTKITHPTEFDFYLCSHAGIQGTSRPSHYHVLW 723

Qy        800 DSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAF 841
              | :    | :| |||: || |      : :|||  ||| :||
Db        724 DDNRFSSDELQILTYQLCHTYVRCTRSVSIPAPAYYAHLVAF 765


RESULT 7
AGOL_ARATH
ID   AGOL_ARATH     STANDARD;      PRT;   997 AA.
AC   Q9SJK3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Argonaute-like protein At2g27880.
GN   AT2G27880 OR T1E2.20.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA   Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA   Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA   Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA   Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA   Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
```

```
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AC006929; AAD21514.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
FT   DOMAIN      359    471       PAZ.
FT   DOMAIN      638    958       PIWI.
SQ   SEQUENCE   997 AA;  111134 MW;  DA93A1446C422F31 CRC64;

  Query Match             10.2%;  Score 460;  DB 1;  Length 997;
  Best Local Similarity   22.3%;  Pred. No. 4.9e-26;
  Matches  219;  Conservative  154;  Mismatches  425;  Indels  184;  Gaps   37;

Qy          3 GRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQS--------------------PT 42
              | :| |  ||    |   |  |  |   ||.     |||
Db         11 GASRGRGGGRRSDQRQDQSSGQVA--WPGL-----QQSYGGRGGSVSAGRGRGNVGRGEN 63

Qy         43 EGDL----------VGRGRQRGMV------VGATSKSQELQISAGFQELSLAERGGRRRD 86
               |||          |  || || :      | ::||:  :  |:  :|    |      :
Db         64 TGDLTATQVPVASAVSGGRGRGNIGDPTFSVASSSKTVSV-ASSSKEESKNTEVSETMSN 122

Qy         87 FHDXGVNTRQNLDHVKESKT----------GSSGIIVKLSTNHFRLTSRPQWALYQYHI 135
                        |:  :  :  : :          |: |  | :  ||| |       || | :
Db        123 LQITSTETKPEMTSLPPASSKAVTFPVRPGRGTLGKKVMVRANHF-LVQVADRDLYHYDV 181

Qy        136 DYNPLMEA----RRLRSALLFQHED--LIGRCHAFD--------GTILFLPKRLQHKVTE 181
                || : :    | :   |:  ::|  | |:  |:|        | : |  |     : |
Db        182 SINPEVISKTVNRNVMKLLVKNYKDSHLGGKSPAYDGRKSLYTAGPLPFDSKEFVVNLAE 241

Qy        182 VFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRR-------LLKIMNL-------- 226
                :  :|:     : : |    ||    |       |:        ::::::
Db        242 KRADGSSGKDRPFKVAVKN---VTSTDLYQLQQFLDRKQREAPYDTIQVLDVVLRDKPSN 298

Qy        227 --QQIGRNYYNPSDPIDIPNHR------LVIWPGFTTSILQYENNIMLCTDVSHKVLRSE 278
                  :||:::: |   |  : |       :  | |:  |:   :  : |  ||| :
Db        299 DYVSVGRSFFHTSLGKDARDGRGELGDGIEYWRGYFQSLRLTQMGLSLNIDVSARSFYEP 358

Qy        279 TVL-DFM---FNLYQQTEEHKFQE--QVSKELIGL-IVLTKYN-NKTYRVDDIDWDQNPK 330
               |: ||:    |:        :  :  :| | |  | : |  :|  |: :: |      :
Db        359 IVVTDFISKFLNIRDLNRPLRDSDRLKVKKVLRTLKVKLLHWNCTKSAKISGISSLPIRE 418

Qy        331 STFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMXIPELC-- 388
                |   | || : ::|: ::||  :     | : :          |:   |     |||
Db        419 LRFTLEDKSEKTVVQYFAEKYNYRVKYQALPAIQT---------GSDTRPVYLPMELCQI 469

Qy        389 -----YLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWG 443
                    |  | :|       :  |   |  |  |:  :  |:   : :|:: :|   :|
Db        470 DEGQRYTKRLNEKQ------VTALLKATCQRPPDRENSIKNLVVKNNYNDDLSKE---FG 520

Qy        444 LSFDSNLLSFSGRILQSE--KIHQGGKTFDYNPQFADWS----KETRGAPLISVKPLDNW 497
              :|  : | |   |:|     | |  ||    ||:   |:     |   || : |      |
Db        521 MSVTTQLASIEARVLPPPMLKYHDSGKEKMVNPRLGQWNMIDKKMVNGAKVTS------W 574

Qy        498 LLI--YTRRNYEAANSLIQNLFKVTPAMGIQMKKA-----IMIEVDDRTEAYLRALQQKV 550
                :  || :       : |  :  : |:: |       |    :   || |     |
Db        575 TCVSFSTRIDRGLPQEFCKQLIGMCVSKGMEFKPQPAIPFISCPPEHIEEALLDI--HKR 632

Qy        551 TSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMN 610
                  |::: :|  :    |  ||:   |:    |||   | : |          :||::|

Db        633 APGLQLLIVIL-PDVTGSYGKIKRICETELGIVSQCCQPRQVNKLN--KQYMENVALKIN 689

Qy        611 CKMGGELWRVDMALKL---------AMIVGIDCYHDTTA--GRRSIAGFVASIN-EGMTR 658
                | ||    :: |::           :|:| |  |          |||  |||::   : :
Db        690 VKTGGRNTVLNDAIRRNIPLITDRPTIIMGADVTHPQPGEDSSPSIAAVVASMDWPEINK 749

Qy        659 WFSRCVFQDRGQELVDGLKVCLQAALR-------------AWSGCNEYMPSRVIVYRDGV 705
                :       |   :|::  |   :|   |              |:      :| |:| |||||
Db        750 YRGLVSAQAHREEIIQDLYKLVQDPQRGLVHSGLIREHFIAFRRATGQIPQRIIFYRDGV 809

Qy        706 GDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGG------RLQN 759
                 :||   :: :|:         |:   | ||:| ::|:|| : | | :  |      :  |
Db        810 SEGQFSQVLLHEMTAIRKACNSLQENYVPRVTFVIVQKRHHTRLFPEQHGNRDMTDKSGN 869

Qy        760 PLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHV 819
                |||||:| ::  |  :||:: | |   |:  | ||:|: | :|   | :| ||   |:
Db        870 IQPGTVVDTKICHPNEFDFYLNSHAGIQGTSRPAHYHVLLDENGFTADQLQMLTNNLCYT 929

Qy        820 YYNWPGVIRVPAPCQYAHKLAF 841
                |      : :  |  |||  ||
Db        930 YARCTKSVSIVPPAYYAHLAAF 951


RESULT 8
YO43_CAEEL
ID   YO43_CAEEL     STANDARD;      PRT;  1040 AA.
AC   P34681;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical 115.4 kDa protein ZK757.3 in chromosome III.
GN   ZK757.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA   Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA   Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA   Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA   Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA   Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA   Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z30215; CAA82941.1; -.
DR   EMBL; Z29121; CAA82941.1; JOINED.
DR   EMBL; Z29121; CAA82389.1; -.
DR   EMBL; Z30215; CAA82389.1; JOINED.
DR   PIR; S41013; S41013.
DR   WormPep; ZK757.3; CE01117.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
```

```
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Hypothetical protein.
FT   DOMAIN      378    486       PAZ.
FT   DOMAIN      660    966       PIWI.
SQ   SEQUENCE   1040 AA;  115415 MW;  4C8483C1F1D72338 CRC64;

  Query Match             9.3%;  Score 423;  DB 1;  Length 1040;
  Best Local Similarity   20.7%;  Pred. No. 2.9e-23;
  Matches  193;  Conservative  153;  Mismatches  357;  Indels  228;  Gaps   34;

Qy        107 GSSGIIVKLSTNHFRL-TSRPQWALYQYHID-YNPLMEARRL---RSALLF-----QHED 156
Db         93 GTIGRQIPVKSNFFAMDLKNPKMVVIQYHVEIHHP--GCRKLDKDEMRIIFWKAVSDHPN 150

Qy        157 LIGR--CHAFDGT-ILFLPKRLQ--------HKVTEVFSQTRNGEHVRITI--------- 196
Db        151 IFHNKFALAYDGAHQLYTVARLEFPDDQGSVRLDCEASLPKDNRDRTRCAISIQNVGPVL 210

Qy        197 -----TLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYN----------PSDPI 240
Db        211 LEMQRTRTNNLDERVLTPIQILDIICRQSLTCPLLKN-SANFYTWKSSCYRIPTAAGQAL 269

Qy        241 DIPNHRLVIWPGFTTSILQYEN-NIMLCTDVSHKVLRSE--TVLDFMFN---------- 286
Db        270 DLEGGK-EMWTGFFSSAHIASNYRPLLNIDVAHTAFYKTRITVLQFMCDVLNERTSKPNR 328

Qy        287 ----------------------------------------------------LYQQT 291
Db        329 NNPRGPGGPGGPGGYRGGRGGGRGGSYGNFGNRGPPGANVRDDFGGNGLTFTMDTLSRDT 388

Qy        292 EEHKFQEQVSKELI-GLIVLTKYNNKTYRVDDIDWDQNP--KSTFKKAD--GSEV--SFL 344
Db        389 QLSSFETRIFGDAIRGMKIRAAHRPNAIRVYKVNSLQLPADKLMFQGIDEEGRQVVCSVA 448

Qy        345 EYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVM 404
Db        449 DYFSEKYG----PLKYPKL---PCLHVGP----------PTRNIFLPM 479

Qy        405 KDLAVHTRLTPEQRQREVGR----------LIDYIHKDDNVQR----------ELRDWG 443
Db        480 EHCLID---SPQKYNKKMSEKQTSAIIKAAAVDATQREDRIKQLAAQASFGTDPFLKEFG 536

Qy        444 LSFDSNLLSFSGRILQSEKIHQGGKTFDYN----PQFADWSKETRGAPLISVKPLDNWLL 499
Db        537 VAVSSQMIQTTARVIQPPPIMFGGNNRSVNPVVFPKDGSWTMDNQTLYMPATCRSYSMIA 596

Qy        500 IYTRRNYEAANSLIQNLFKVTPAMGIQMKK----------AIMIEVDDRTEAYLR 544
Db        597 LVDPRDQTSLQTFCQSLTMKATAMGMNFPRWPDLVKYGRSKEDVCTLFTEIAD------- 649

Qy        545 ALQQKVTSDT-QIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIAT 603
Db        650 --EYRVTNTVCDCIIVVLQSKNSDIYMTVKEQSDIVHGIMSQCVLMKNVSR--PTPATCA 705

Qy        604 KIALQMNCKMGGGELWRVDMALKL---------AMIVGIDCYHDTTAGRR----SIAGFVA 650
Db        706 NIVLKLNMKMGGINSRI-VADKITNKYLVDQPTMVVGIDVTHPTQAEMRMNMPSVAAIVA 764

Qy        651 SINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQL 710
Db        765 NVDLLPQSYGANVKVQKKCRESVVYLLDAIRERIITFYRHTKQKPARIIVYRDGVSEGQF 824

Qy        711 KTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQ----SGGRLQNPLPGTVI 766
Db        825 SEVLREEIQSIRTACLAIAEDFRPPITYIVVQKRHHARIFCKYQNDMVGKAKNVPPGTTV 884

Qy        767 DVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGV 826
Db        885 DTGIVSPEGFDFYLCSHYGVQGTSRPARYHVLLDECKFTADEIQSITYGMCHTYGRCTRS 944

Qy        827 IRVPAPCQYAHKLAFLVGQSIHREPNLSLSN 857
Db        945 VSIPTPVYYADLVATRARCHVKRKLGLADNN 975


RESULT 9
I2C2_HUMAN
ID   I2C2_HUMAN     STANDARD;      PRT;   377 AA.
AC   Q9UKV8;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2)
DE   (Fragment).
GN   EIF2C2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20005943; PubMed=10534406;
RA   Koesters R., Adams V., Betts D., Moos R., Schmid M., Siermann A.,
RA   Hassam S., Weitz S., Lichter P., Heitz P.U., von Knebel Doeberitz M.,
RA   Briner J.;
RT   "Human eukaryotic initiation factor EIF2C1 gene: cDNA sequence,
RT   genomic organization, localization to chromosomal bands 1p34-p35, and
RT   expression.";
RL   Genomics 61:210-218(1999).
CC   -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC       initiation process.
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF121255; AAF13034.2; -.
DR   Genew; HGNC:3263; EIF2C2.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Initiation factor; Protein biosynthesis.
FT   NON_TER       1      1
FT   DOMAIN       35    336       PIWI.
SQ   SEQUENCE   377 AA;  42529 MW;  3BBCE91689D58526 CRC64;

  Query Match             7.8%;  Score 355;  DB 1;  Length 377;
  Best Local Similarity   28.4%;  Pred. No. 7e-19;
  Matches   90;  Conservative   66;  Mismatches  145;  Indels   16;  Gaps    7;

Qy        537 DRTEAYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQ 596
Db         17 DSVEPMFRHL-KNTYAGLQLVVVIL-PGKTPVYAEVKRVGDTVLGMATQCVQMKNV--QR 72

Qy        597 TVMAIATKIALQMNCKMGG------ELWRVDMALKLAMIVGIDCYHDTTA-GRR-SIAGF 648
Db         73 TTPQTLSNLWLKINVKLGGVNNILLPQGRPPVFQEPVIFLGADVTHPPAGDGKKPSIAAV 132

Qy        649 VASINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDG 708
Db        133 VGRMDAHPNRYCATVRVQQHRQEIIQDLAAMVRELLIQFYKSTRFKPTRIIFYRDGVSEG 192

Qy        709 QLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFA----QSGGRLQNPLPGT 764
Db        193 QFQQVLHHELLAIREACIKLEKDYQPGITFIVVQKRHHTRLFCTDKNERVGKSGNIPAGT 252

Qy        765 VIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWP 824
```

```
                :| ::| |  :||:: | |   |:  |:||:|::| :    | :| |||: || |
Db        253 TVDTKITHPTEFDFYLCSHAGIQGTSRPSHYHVLWDDNRFSSDELQILTYQLCHTYVRCT 312

Qy        825 GVIRVPAPCQYAHKLAF 841
              : :|||  ||| :||
Db        313 RSVSIPAPAYYAHLVAF 329


RESULT 10
YQ53_CAEEL
ID   YQ53_CAEEL     STANDARD;      PRT;  1032 AA.
AC   Q09249;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein C16C10.3 in chromosome III.
GN   C16C10.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Lloyd C.;
RL   Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: STRONG, TO C.ELEGANS C14B1.7. ALSO TO C06A1.4,
CC       F48F7.1, R06C7.1, T22B3.2 AND ZK757.3.
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z46787; CAA86741.1; -.
DR   WormPep; C16C10.3; CE01494.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Hypothetical protein.
FT   DOMAIN      369    481       PAZ.
FT   DOMAIN      650    977       PIWI.
SQ   SEQUENCE   1032 AA;  117105 MW;  EBEAE23CE67BB149 CRC64;

  Query Match             5.2%;  Score 237;  DB 1;  Length 1032;
  Best Local Similarity   21.5%;  Pred. No. 1.8e-09;
  Matches  134;  Conservative  100;  Mismatches  254;  Indels  136;  Gaps   26;

Qy        293 EHKFQEQ------VSKELIGL-IVLTKYNNKTYRVDDIDWDQNPKSTFKKADGS---EVS 342
              || | |:      : : | || : |       ||  |  :    ::|   |     ||:
Db        381 EHAFGERGNTKRCIEEALKGLDVECTHLKGNLIRVSSIAENNAENTSFMMKDDKGEREVT 440

Qy        343 FLEYYRKQYNQEITDLKQPVLVSQPKR----------RRGPGGTLPGPAMXIPELCYLTG 392
                ||:  ||| ::   : |::||:  :          |  ||  :    |       :||
Db        441 VAEYFLLQYNIKLKYPRLPLVVSKRFKHESFFPMELLRIAPGQRIKVNKMSPTVQSAMTG 500

Qy        393 LTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLLS 452
                |                |:   : |  ::      ||:: |   :  :|   |:|
Db        501 RNASM----------------PQHHVKLVQDIL-----RDNLKLEQNKYMDAFGIKLMS 538

Qy        453 -----FSGRILQSEKIHQGGKTF--DYN-PQFADWSKETRGAPL--ISVKPLDNWLLIYT 502
                    : ::|   :|   |:|:  | : | |    |    | :  | :   || :
Db        539 TEPIQMTAKLLPPAQIKFKGQTYMPDMSRPAFRTQDKFVEPARIRKIGIVVFDNCI---- 594

Qy        503 RRNYEAANSLIQNLFKVTPAMGIQMKKA----IMIEVDDRTEAYLRALQQKVTSD-TQIV 557
```

```
                |      |      || ::|      : |::      :: | :|   |   |:
Db        595 --QMRQAEDFCDKLSNFCRDNGITVEKDSRDWSIRELNSSDSVAIQNLMKKWLDDRVDIL 652

Qy        558 VCLLSSNRKDKYDAIKKY----------LCTDCPTPSQCVVARTLGKQ---QTVMAIATK 604
              | :    : | :| :| :          ||       |  | : :| |   ||:  :  |
Db        653 VGIAREKKPDVHDILKYFEESIGLQTIQLC-------QQTVDKMMGGQGGRQTIDNVMRK 705

Qy        605 IALQMNCKMGGELWRVDM--ALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWFSR 662
                | | ||  : |::  |::   :    | :: |  |: :   |  |   ::   ||
Db        706 ----FNLKCGGTNFFVEIPNAVRGKAV----CSNNETL-RKKLLEHVQFIGFEISHGASR 756

Qy        663 CVFQDRGQELVDG--------------------------------LKVCLQAALRAWSG 689
              :| || :  :||                                |      :|::
Db        757 TLF-DRSRSQMDGEPSVVGVSYSLTNSTQLGGFTYLQTQKEYKLQKLDEFFPKCVRSYKE 815

Qy        690 CNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARF 749
              :: :|:|:::|| | |:|     |  || :       :  || | | ||: ::  :||
Db        816 HSKTLPTRIVIYRVGAGEGNFNR-VKEEVEEMRRTFDKIQPGYRPHLVVIIAQRASHARV 874

Qy        750 FAQ--SGGRL--QNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLK 805
              |    || |   ||   || ::  :|   : :| : ||    |:| |  | :: : :
Db        875 FPSCISGNRATDQNIPSGTCVENVLTSYGYDEFILSSQTPLIGTVRPCKYTILVNDAKWS 934

Qy        806 PDHIQRLTYKXC---HVYYNWPGV 826
              : :  |||       | |  | |
Db        935 KNELMHLTYFRAFGHQVSYQPPSV 958


RESULT 11
RPOD_SINAL
ID   RPOD_SINAL     STANDARD;      PRT;  1384 AA.
AC   Q9THV5;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA-directed RNA polymerase beta" chain (EC 2.7.7.6).
GN   RPOC2.
OS   Sinapis alba (White mustard) (Brassica hirta).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Sinapis.
OX   NCBI_TaxID=3728;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 6-20.
RC   STRAIN=cv. Albatros; TISSUE=Cotyledon;
RX   MEDLINE=20069369; PubMed=10601874;
RA   Pfannschmidt T., Ogrzewalla K., Baginsky S., Sickmann A., Meyer H.E.,
RA   Link G.;
RT   "The multisubunit chloroplast RNA polymerase A from mustard (Sinapis
RT   alba L.). Integration of a prokaryotic core into a larger complex
RT   with organelle-specific functions.";
RL   Eur. J. Biochem. 267:253-261(2000).
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC       OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC       SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: IN CHLOROPLAST THE RNA POLYMERASE IS COMPOSED OF FOUR
CC       SUBUNITS: ALPHA, BETA, BETA', AND BETA".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ243754; CAB48415.2; -.
DR   HSSP; Q9KWU6; 1HQM.
DR   InterPro; IPR000722; RNA_pol_A.
DR   InterPro; IPR002879; RNA_pol_A2.
```

```
DR   Pfam; PF00623; RNA_pol_A; 1.
DR   Pfam; PF01854; RNA_pol_A2; 1.
KW   Transferase; Transcription; DNA-directed RNA polymerase; Chloroplast.
SQ   SEQUENCE   1384 AA;  157740 MW;  9A7B4D02053F4A80 CRC64;

  Query Match             3.1%;  Score 142.5;  DB 1;  Length 1384;
  Best Local Similarity   20.1%;  Pred. No. 0.029;
  Matches  132;  Conservative   99;  Mismatches  201;  Indels  225;  Gaps   38;

Qy         39 QSPTEGDLVGRGRQRGMVVGATSKSQELQISAGFQELSLAERGGR--RRDFHDXGVNTRQ 96
              :||| |||| |   |:: |                  |: | | :   | ||  || |
Db        310 RSPTHGDLVELGEAVGIIAGQ---------------SIGEPGTQLTLRTFHTGGVFTGG 353

Qy         97 NLDHVKESKTG----SSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRL------ 146
                :||:    |    :  ::    | |      | :  |   || : ::|:  :
Db        354 TAEHVRAPYNGKIKFNEDLVHSTRTRH----GHPAFLCY---IDLSVIIESEDIIHSVTI 406

Qy        147 --RSALLFQHEDLIGR----CHAFDGTILFLPKRLQHKVTEVFSQTRNGEH--------V 192
                :| || |::  :          :|| |  |    |   ::| :  | |
Db        407 PPKSFLLVQNDQYVESEQVIAEIREGTYTFQFKERVRKY--IYSDSEGGMHWSTDVSHAP 464

Qy        193 RITITLTNELPPTS-------PTC-----------------LQFYNIXFRRLLKI-MNLQ 227
                | :  : || ||       :|                 : | ::  : :  : :|
Db        465 EFTYSNVHLLPKTSHLWILSGGSCGSSLILFSIHKDQDQMNIPFLSVERKSISSLSVNND 524

Qy        228 QIGRNYYNPSDPID----IPNHRLV------IWPGFTTSILQYENNIMLCTDVSHKVLRS 277
              |: : ::: ||  |    |||:  :           |  | : :||     :|:  |  |:
Db        525 QVSQKFFS-SDFSDKKILIPNYSELNGIVGTSHYNFIYSAIFHEN-----SDLLAKRRRN 578

Qy        278 ETVLDFMFNLYQQTEEHKF--QEQVSKE----------LIGLIVLTKYNNK-------- 316
                :: |  :: :| :| :|     :| |           :       :|  |
Db        579 RFLIPFQ-SIQEQEQEKEFIPHSGISVEIPINGIFRRNSIFAFFDDPRYRRKSSGILKYG 637

Qy        317 TYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGT 376
              | : | |          :| | |   :: :: :|  ::
Db        638 TLKADSI---------IQKEDMIEYRGVQKFKTKYEMKVDRF---------------FF 672

Qy        377 LPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQ 436
              :|     :||   :      |  :::   : | ||:|   |  :|| ||    |   ::
Db        673 IPEEVHILPESSAI------MVENYSI---IGVDTRITLNIRS-QVGGLIRVERKKKRIE 722

Qy        437 RELRDWGLSFDSNLLSFSGRIL---QSEKI--HQG-------GKTFDYNPQFADWSKETR 484
                ::          ||| |    :::||   | |       |||          |||:
Db        723 LKI------------FSGDIHFPDKTDKISRHSGILIPPGRGKTN---------SKES- 759

Qy        485 GAPLISVKPLDNWLLIYTRR----------------NYEAANSLIQNLFKVTPAMGIQMK 528
                  | : ||  || :|                  || |:|:  ||  : |    : |
Db        760 -------KIVKNW--IYVQRITPTKKKFFVLVRPVATYEIADSI--NLATLFPKDLFREK 808

Qy        529 KAIMIEVDDRTEAYLRALQQKVT---SDTQIVV---CLL----SSNRKDKYDAIKKY 575
                | : | :    |:     | |   ||| | :   ||:      |:    : :: :
Db        809 DNIQLRVFN----YILYGNGKPTRGISDTSIQLVRTCLVLNWDQDNKNSSLEEVRAF 861


RESULT 12
NXF1_MOUSE
ID   NXF1_MOUSE     STANDARD;      PRT;   618 AA.
AC   Q99JX7; O88985;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Nuclear RNA export factor 1 (Tip associating protein) (Tip-associated
DE   protein) (mRNA export factor TAP).
GN   NXF1 OR TAP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart ventricle;
RA   Tannoch V.J., Cormier-Regard S., Claycomb W.C.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Involved in the nuclear export of mRNA species bearing
CC       retroviral constitutive transport elements (CTE) and in the export
CC       of mRNA from the nucleus to the cytoplasm (By similarity).
CC   -!- SUBUNIT: Interacts with NXT1, NXT2, E1B-AP5, RAE1, the REF
CC       proteins, and with several nucleoporins (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear; Localized predominantly in the
CC       nucleoplasm and at both the nucleoplasmic and cytoplasmic faces of
CC       the nuclear pore complex. Shuttles between the nucleus and the
CC       cytoplasm (By similarity).
CC   -!- TISSUE SPECIFICITY: Expressed ubiquitously.
CC   -!- DOMAIN: The minimal CTE binding domain consists of an RNP-type RNA
CC       binding domain (RBD) and leucine-rich repeats (By similarity).
CC   -!- DOMAIN: The nucleoporin binding domain consists of a NTF2-like
CC       domain and a UBA-like domain. The NTF2 domain heterodimerizes with
CC       NXT1 AND NXT2. The formation of NXF1/NXT1 heterodimers is required
CC       for NXF1-mediated nuclear mRNA export. The UBA-like domain
CC       mediates direct interactions with nucleoporin-FG-repeats and is
CC       necessary and sufficient for localization of NXF1 to the nuclear
CC       rim. The conserved loop 594-NWD-596 of the UBA domain has a
CC       critical role in the interaction with nucleoporins (By
CC       similarity).
CC   -!- DOMAIN: The leucine-rich repeats and the NTF2-domain are essential
CC       for the export of mRNA from the nucleus (By similarity).
CC   -!- MISCELLANEOUS: The RNA-binding domain is a non-canonical RNP-type
CC       domain.
CC   -!- SIMILARITY: BELONGS TO THE NXF FAMILY.
CC   -!- SIMILARITY: CONTAINS 4 LEUCINE-RICH REPEATS (LRR).
CC   -!- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC   -!- SIMILARITY: CONTAINS 1 UBA DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF093140; AAC63368.1; -.
DR   EMBL; BC005594; AAH05594.1; -.
DR   MGD; MGI:1858330; Nxf1.
DR   InterPro; IPR001611; LRR.
DR   Pfam; PF00560; LRR; 2.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00446; LRRcap; 1.
DR   PROSITE; PS50177; NTF2_DOMAIN; 1.
KW   Transport; mRNA transport; Nuclear protein; RNA-binding; Repeat;
KW   Leucine-rich repeat; Multigene family.
FT   DOMAIN      118    197       RNA-BINDING (RRM) (BY SIMILARITY).
FT   REPEAT      265    290       LRR 1.
FT   REPEAT      291    314       LRR 2.
FT   REPEAT      315    342       LRR 3.
FT   REPEAT      343    370       LRR 4.
FT   DOMAIN      385    535       NTF2.
FT   DOMAIN      565    607       UBA-LIKE (BY SIMILARITY).
FT   DOMAIN       66     99       NUCLEAR LOCALIZATION SIGNAL (BY
FT                                SIMILARITY).
FT   DOMAIN       82    109       NUCLEAR EXPORT SIGNAL (BY SIMILARITY).
FT   DOMAIN      550    560       PRO-RICH.
FT   CONFLICT     79     79       D -> N (IN REF. 2).
SQ   SEQUENCE   618 AA;  70300 MW;  0556EF864C935435 CRC64;

  Query Match             2.7%;  Score 122.5;  DB 1;  Length 618;
  Best Local Similarity   20.4%;  Pred. No. 0.26;
  Matches  132;  Conservative   83;  Mismatches  248;  Indels  185;  Gaps   31;

Qy         48 GRGRQRGMVVGATSKSQELQISAGFQELSLAERGGR--------RRDFHDXGVNTRQNL- 98
                | | :|    |:  :|   :   |   ::  : | |        | |    | : |  :
```

```
Db         33 GEGNRRSGRGGSGVQSSRFEEDDGDVAMNDPQDGPRVRYNPYTNRPDRRGDGWHDRDRIH 92

Qy         99 -----DHVKESKTGSSGIIVKLSTNHFRLT------SRPQWALYQYH----IDYNPLMEA 143
                    |     : |:       : | |::|          | |        : :|||:
Db         93 ITVRRDRAPAERGGAGTSQDGTTKNWFKITIPYGRKYDKTWLLSMIQSKCSVPFNPIE-- 150

Qy        144 RRLRSALLFQHEDLIGRCHAF--DGTILFLPKRLQHKVTEVFSQTRNGEHVRITI----- 196
                   | :|:   | | |  | |     | : ||:     | |    : | |
Db        151 --------FHYENT--RAHFFVEDATTASALKGVNHKI-----QDRENRRISIIINASAP 195

Qy        197 --TLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNP-----SDP------IDIP 243
                |: ||| |              :|  ||: :  |          |||      ||:
Db        196 PYTVQNELKPEQ----------IEQLKLIMSKRYDGNQQALDLKGLRSDPDLVAQNIDVV 245

Qy        244 NHRLVIWPGFTTSILQYENNI--MLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQVS 301
                :|         ::   | ||  :|  ::|   |     || | :: |:    |
Db        246 LNRRSC---MAATLRIIEENIPELLSLNLSSNRLYK---LDDMSSIVQKAPNLKTLNLSG 299

Qy        302 KELIGLIVLTKYNNKTYRVDDIDWDQNPK-STFKKADGSEVSFLEYYRKQYNQEITDLKQ 360
                 ||     | |   |  ::::: |:||    |    | : |::   |:::
Db        300 NELKTERELDKI--KGLKLEELWLDRNPMCDNF----GDQSSYISAIRERF--------- 344

Qy        361 PVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 420
                      ||  |  | || |                :  |:   | | | :
Db        345 ------PKLLRLDGHELPPP---------------------ISFDVEAPTMLPPCK--- 373

Qy        421 EVGRLIDYIHKDDNVQRELRDWGLSFDSN-----LLSFSGRILQSEKIHQG---GKTFDY 472
                  |      :    | | |: : : :||      | ::          | |     :  |
Db        374 --GSYFGTENLKSLVLRFLQQYYVIYDSGDRQGLLYAY----------HDGACCSLSIPY 421

Qy        473 NPQ------FADWSKETRGAPLISVKPLDNW-LLIYTRRNYEA----------ANSLIQ 514
                |||       |:: |::|    :  :|   : || :|| |  |           |: :
Db        422 NPQNPVRKNLAEYVKDSRNVKKLK-EPTQRFRLLKHTRLNVVAFLNELPKTQHDVNAFVV 480

Qy        515 NLFKVTPAMGIQMKKAIMIEVDDRTEAYLRALQQKVTSDTQIVVCLLSSN---------- 564
                ::   |  :       :  ||| ::   |||        |:  : : :||
Db        481 DISAQTSTLLCFSVNGVFKEVDGKSRDSLRAF-------TRTFIAVPASNSGLCIVNDEL 533

Qy        565 --RKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMN 610
                  |    : |::      ||||   |  ||  :|  |  |      ||
Db        534 FVRNASPEEIQRAFAMSAPTPSSSPVP-TLSPEQQEMLQAFSTQSGMN 580


RESULT 13
YN13_YEAST
ID   YN13_YEAST     STANDARD;      PRT;  1238 AA.
AC   P53840;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Hypothetical 141.1 kDa protein in MET2-SEC2 intergenic region.
GN   YNL273W OR N0636.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Messenguy F., Dubois E., Vierendeels F., Scherens B., Pierard A.,
RA   Glansdorff N.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z71549; CAA96181.1; -.
DR   SGD; S0005217; YNL273W.
KW   Hypothetical protein.

SQ   SEQUENCE   1238 AA;  141119 MW;  CF8FEC33EE0088E9 CRC64;

  Query Match             2.7%;  Score 122.5;  DB 1;  Length 1238;
  Best Local Similarity   18.0%;  Pred. No. 0.74;
  Matches  112;  Conservative  113;  Mismatches  234;  Indels  163;  Gaps   25;

Qy         35 PRPQQSPTEGDLVGRGRQRGMVVGATSKSQELQISAGFQELSLAERGGRRRDFHDXGVNT 94
              ||  :: |   :| :          :| :  :  ||||:   | ::  : |
Db        304 PRQFETGTHSKVVNKNE--------SSSANNIVTSAGFELSKLLQKEHQMR--------- 346

Qy         95 RQNLDHVKESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQH 154
              :  : |      :   |::   : :  |||     ||    |    | ::::    :: :|
Db        347 KNVIKHTSARHSRFGGLLSIQTPDKTRLTVSGSQALVDEKIALQKLDDSKKWNKRIIKKH 406

Qy        155 EDLIGRCHAFDGTILFLPKRLQHKVT--EVFSQTRNGEHVRITITLTNELPPTSPTCLQF 212
              : :    | :|     ||  | : . |   :|    ||:| :  |                |
Db        407 QSV-----AAEG----LPNSLLNSQTGKAIFFTESNGKHFKEFIN------------NF 444

Qy        213 YNIXFRRLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVSH 272
              :  |  ||  :       ||:          |:|       :: | |  :     | :
Db        445 IDSGFNILLHSVT------NYFTTE------QDRMV-------TLEQVEYLLFFAWFVKY 485

Qy        273 KVLRSETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTYRVDD---IDWDQ-- 327
              ::|||:  :|   :: | :|  |   :|:  |:  :: : |: | : |     | :::
Db        486 QLLRSK--IDNSADIKQVSEALK---EVTFILVSSLLRSAYDLKNWTVTHAGMIAFNELL 540

Qy        328 NPKSTFKKA---DGSEVSFL-------------------EYYRKQYNQEITDLKQPVL- 363
              |  | | |   | ::: |:                       |  |: :   :|   ||
Db        541 NLVSRTKAAQEEDSTDIEFIVSRLFSDERIQLLSNLPKIGSKYSLQFMKSCIELTHSVLK 600

Qy        364 ------------VSQPKRRRGPGGTLPGPAMXIPE---------LCYLTGLTDKMRNDF- 401
                         :   ||:       |    : |          |  ||      :  :|
Db        601 VLEQYSDDKTLVIEGKSRRQKKFNISEGDITKLIEEENVDRDEALDILTSSLRSIEVNFQ 660

Qy        402 ----NVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQR----------ELRDWGLSFD 447
                  | | :  : | :   :| ||:        :||::::          : ::  | |
Db        661 KVQANYMTEPVIETYINFLERFREL--------EDDSIKKVFSFFHRVFVQAKEQALLFR 712

Qy        448 SNLLSFSGRILQSEKIHQGGKTFDYNPQFADW-----SKETRGAPL----ISVKPLDNWL 498
              :|:     :|  : : :  ::  |  ||:|:      |  : :|     :   || |
Db        713 FDLIILLREMLSPDGLDRMSRSRKYVSQFSDYFLARLKKRLKKSPAWFVGLLFPPLHNSE 772

Qy        499 LIYTRR-------NYEAANSLIQNLFK-------VTPAMGIQMKKAIMIE--VDDRTEAY 542
              : : :|       | |:  :   : ||       : |:: : ||  :::   :||
Db        773 VGFYQRYGEYNVLNNESMYAAPASQFKPIPDEEALPPSILLDMKYGVLVSTLLDDGKTEL 832

Qy        543 LRALQQKVTSDTQIVVCLLSSN 564
              |  | : :|    |    |: |
Db        833 LDQLLKHITHTLDIFKSWLTVN 854


RESULT 14
RPOD_ARATH
ID   RPOD_ARATH     STANDARD;      PRT;  1376 AA.
AC   P56764;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA-directed RNA polymerase beta" chain (EC 2.7.7.6).
GN   RPOC2.
OS   Arabidopsis thaliana (Mouse-ear cress).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20039611; PubMed=10574454;
RA   Sato S., Nakamura Y., Kaneko T., Asamizu E., Tabata S.;
RT   "Complete structure of the chloroplast genome of Arabidopsis
RT   thaliana.";
```

```
RL   DNA Res. 6:283-290(1999).
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC       OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC       SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: IN CHLOROPLAST THE RNA POLYMERASE IS COMPOSED OF FOUR
CC       SUBUNITS: ALPHA, BETA, BETA', AND BETA".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP000423; BAA84375.1; -.
DR   HSSP; Q9KWU6; 1HQM.
DR   InterPro; IPR000722; RNA_pol_A.
DR   InterPro; IPR002879; RNA_pol_A2.
DR   Pfam; PF00623; RNA_pol_A; 1.
DR   Pfam; PF01854; RNA_pol_A2; 2.
KW   Transferase; Transcription; DNA-directed RNA polymerase; Chloroplast.
SQ   SEQUENCE   1376 AA;  156365 MW;  7CBB5820163E2B9D CRC64;

  Query Match             2.7%;  Score 121.5;  DB 1;  Length 1376;
  Best Local Similarity   19.5%;  Pred. No. 1;
  Matches  126;  Conservative  102;  Mismatches  210;  Indels  209;  Gaps   35;

Qy         39 QSPTEGDLVGRGRQRGMVVGATSKSQELQISAGFQELSLAERGGR--RRDFHDXGVNTRQ 96
              :||| ||||  |   |:: |                          |: | | :   | ||  || |
Db        306 RSPTHGDLVELGEAVGIIAGQ---------------SIGEPGTQLTLRTFHTGGVFTGG 349

Qy         97 NLDHVKESKTG----SSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRL------ 146
                :||:    |    :  ::    | |      | :  |   || : ::|:  :
Db        350 TAEHVRAPYNGKIKFNEDLVHPTRTRH----GHPAFLCY---IDLSVIIESEDIIHSVTI 402

Qy        147 --RSALLFQHEDLIGR----CHAFDGTILFLPKRLQHKVTEVFSQTRNGEH--------V 192
                :| || |::  :          :||  |  |     |   ::| :    |
Db        403 PPKSFLLVQNDQYVESEQVIAEIREGTYTFHFKERVRKY--IYSDSEGEMHWSTDVSHAP 460

Qy        193 RITITLTNELPPTS-----------PTCLQF--------YNIXF----RRLLKIM--NLQ 227
                | :  : || ||            : ::|         || |    |: :  :  |
Db        461 EFTYSNVHLLPKTSHLWILSGGSCGSSLIRFSIHKDQDQMNIPFLSAERKSISSLSVNND 520

Qy        228 QIGRNYYNP--SDPIDI------------PNHRLVIWPGFTTSILQYENNIMLCTDVSH 272
              |: : :::   :||  :             :|  :|:     | : :||     :|:
Db        521 QVSQKFFSSDFADPKKLGIYDYSELNGNLGTSHYNLIY-----SAIFHEN-----SDLLA 570

Qy        273 KVLRSETVLDFMFNLYQQTEEHKFQEQVSKE----------LIGLIVLTKYNNK----- 316
              |  |:  :: |  :: :| :|    |  :| |          :       :|  |
Db        571 KRRRNRFLIPFQ-SIQEQEKEFIPQSGISVEIPINGIFRRNSIFAFFDDPRYRRKSSGIL 629

Qy        317 ---TYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGP 373
                 | : | |          :| |  |   ::  : :|  ::
Db        630 KYGTLKADSI---------IQKEDMIEYRGVQKIKTKYEMKVDRF--------------- 665

Qy        374 GGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDD 433
                :|     :||   :     |  ::::   : | ||||   |  :|| ||    |
Db        666 -FFIPEEVHILPESSAI------MVQNYSI---IGVDTRLTLNIRS-QVGGLIRVEKKKK 714

Qy        434 NVQRELRDWGLSFDSNLLSFSGRIL---QSEKIH----------QGGKTFDYNPQFADWS 480
               :: ::             ||| |    :::||           :| |    : :| :|
Db        715 RIELKI------------FSGDIHFPDKTDKISRHSGILIPPGRGKKNSKESKKFKNWI 761

Qy        481 KETRGAP-----LISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEV 535
                |  |       : |:|:           || |:|:  ||  : |    : |  | : |
Db        762 YVQRITPTKKKFFVLVRPV---------ATYEIADSI--NLATLFPQDLFREKDNIQLRV 810

Qy        536 DDRTEAYLRALQQKVT---SDTQIVV---CL-LSSNRKDKYDAIKKY 575
                :    |:    | |   ||| | :   || |: ::    : :: :
Db        811 FN----YILYGNGKPTRGISDTSIQLVRTCLVLNWDKNSSLEEVRAF 853
```

```
RESULT 15
PMA6_ARATH
ID   PMA6_ARATH     STANDARD;      PRT;   949 AA.
AC   Q9SH76;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   ATPase 6, plasma membrane-type (EC 3.6.3.6) (Proton pump 6).
GN   AHA6 OR AT2G07560 OR F9A16.7.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J., Ronning C.M., Koo H.L., Moffat K.S.,
RA   Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L., Tallon L.J.,
RA   Gill J.E., Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M.,
RA   Somerville C.R., Copenhaver G.P., Preuss D., Nierman W.C., White O.,
RA   Eisen J.A., Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
CC   -!- FUNCTION: THE PLASMA MEMBRANE H+ ATPASE OF PLANTS AND FUNGI
CC       GENERATES A PROTON GRADIENT THAT DRIVES THE ACTIVE TRANSPORT OF
CC       NUTRIENTS BY H+-SYMPORT. THE RESULTING EXTERNAL ACIDIFICATION
CC       AND/OR INTERNAL ALKINIZATION MAY MEDIATE GROWTH RESPONSES (BY
CC       SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +
CC       H(+)(Out).
CC   -!- SUBUNIT: BINDS TO 14-3-3 PROTEINS. THE BINDING IS INDUCED BY
CC       PHOSPHORYLATION OF THR-948. BINDING TO 14-3-3 PROTEINS ACTIVATES
CC       THE H+-ATPASE (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC       (E1-E2 ATPASES). SUBFAMILY IIIA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AC007662; AAD32758.1; -.
DR   InterPro; IPR001757; ATPase_E1-E2.
DR   InterPro; IPR004014; Cation_ATPase.
DR   InterPro; IPR000695; H_ATPase.
DR   InterPro; IPR001454; Hlgnase/hydrlase.
DR   Pfam; PF00122; E1-E2_ATPase; 1.
DR   Pfam; PF00690; Cation_ATPase_N; 1.
DR   Pfam; PF00702; Hydrolase; 1.
DR   PRINTS; PR00119; CATATPASE.
DR   PRINTS; PR00120; HATPASE.
DR   PROSITE; PS00154; ATPASE_E1_E2; 1.
KW   Hydrolase; Hydrogen ion transport; Transmembrane; Phosphorylation;
KW   ATP-binding; Metal-binding; Magnesium; Multigene family.
FT   DOMAIN        1     64       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     65     84       1 (POTENTIAL).
FT   DOMAIN       85     96       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     97    117       2 (POTENTIAL).
FT   DOMAIN      118    246       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    247    267       3 (POTENTIAL).
FT   DOMAIN      268    276       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    277    294       4 (POTENTIAL).
```

```
FT   DOMAIN      295    645       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    646    667       5 (POTENTIAL).
FT   DOMAIN      668    672       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    673    695       6 (POTENTIAL).
FT   DOMAIN      696    711       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    712    732       7 (POTENTIAL).
FT   DOMAIN      733    753       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    754    774       8 (POTENTIAL).
FT   DOMAIN      775    786       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    787    807       9 (POTENTIAL).
FT   DOMAIN      808    815       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    816    836       10 (POTENTIAL).
FT   DOMAIN      837    949       CYTOPLASMIC (POTENTIAL).
FT   MOD_RES     332    332       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     948    948       PHOSPHORYLATION (BY SIMILARITY).
FT   METAL       590    590       MAGNESIUM (BY SIMILARITY).
FT   METAL       594    594       MAGNESIUM (BY SIMILARITY).
FT   SITE        947    949       INTERACTION WITH 14-3-3 PROTEINS (BY
FT                                SIMILARITY).
SQ   SEQUENCE   949 AA;  105011 MW;  25C076607DB214EF CRC64;

  Query Match             2.7%;  Score 121;  DB 1;  Length 949;
  Best Local Similarity   19.6%;  Pred. No. 0.64;
  Matches  140;  Conservative  108;  Mismatches  277;  Indels  190;  Gaps   37;

Qy         43 EGDLVGRGRQRGMVVGATSKSQELQ-----------------ISAGFQELSLAERGGRR 84
              ||      || |  : ||    ::::                 :|    : ||  |||
Db         34 EGLSSEEGRNRLQIFGANKLEEKVENKFLKFLGFMWNPLSWVMEAAAIMAIVLANGGGRP 93

Qy         85 RDFHD-XGVN----TRQNLDHVKESKTG--SSGIIVKLSTNHFRLTSRPQWALYQYHI-- 135
               |: |  |:         :  ::|:  |  :: ::  |:    ::    :|   :  |
Db         94 PDWQDFVGITCLLIINSTISFIEENNAGNAAAALMANLAPK-TKVLRDGRWGEQEAAILV 152

Qy        136 ----------DYNPLMEARRLR-SALLFQHEDLIGRCHAFDGTILFLPKRLQHKVTEVFS 184
                       |  |  :|| |    |       | |           ||   :|:  ||||
Db        153 PGDLISIKLGDIVP-ADARLLEGDPLKIDQSALTGES---------LP-ATKHQGDEVFS 201

Qy        185 QT--RNGEHVRITITL---------------TNE-------LPPTSPTCLQFYNIXFRRL 220
               :  : ||   : |                ||       |      |:   :|    |
Db        202 GSTCKQGEIEAVVIATGVHTFFGKAAHLVDSTNNVGHFQKVLTAIGNFCI--CSIGIGML 259

Qy        221 LKIMNLQQI-GRNYYNPSD----------PIDIP----------NHRLVIWPGFT---TS 256
              ::|: :  |  | | :  |          || :|          :|||      |   |:
Db        260 IEIIIMYPIQHRKYRDGIDNLLVLLIGGIPIAMPTVLSVTMAIGSHRLSQQGAITKRMTA 319

Qy        257 ILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNK 316
              | :     :||:| :  :  ::  :|   || :      | : | |: : |:|:   ::
Db        320 IEEMAGMDVLCSDKTGTLTLNKLTVD--KNLIE-----VFSKDVDKDYV--ILLSARASR 370

Qy        317 TYRVDDIDWD-----QNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRR 371
                | ||        :||    :|  :|| || :   :    ||       : :  :  |
Db        371 VENQDAIDTSIVNMLGDPKEA--RAGITEVHFLPFNPVEKRTAIT-----YIDTNGEWHR 423

Qy        372 GPGGTLPGPAMXIPELCYLTGLTDKMRNDF------NVMKDLAVHTRLTPEQRQREVGRL 425
                |    |   | ||| | | | :  ::         :: | |  :  ||: :   |
Db        424 CSKG---APEQII-ELCDLKGETKRRAHEIIDKFAERGLRSLGVARQRVPEKDKESAGTP 479

Qy        426 IDYI----------HKDDNVQRELRDWGLSFDSNLLSFSGRIL-----QSEKIHQGGKTF 470
              :::          |      |   | |:    |:   :|  |        ::  |   :
Db        480 WEFVGLLPLFDPPRHDSAETIRRALDLGV----NVKMITGDQLAIGKETGRRLGMGTNMY 535

Qy        471 DYNPQFADWSKETRGAPLIS-VKPLDNWLLIYTRRNYEAANSLIQNLFKV---------T 520
                :    :    | | |:   ::  | :  ::    ||    | :    |
Db        536 PSSSLLENKDDTTGGVPVDELIEKADGFAGVFPEHKYEIVRKLQERKHIVGMTGDGVNDA 595

Qy        521 PAMGIQMKKA-IMIEVDDRTEAYLRALQQKVTSD--TQIVVCLLSSNRKDKYDAIKKYLC 577
              ||    :||| | | ||| |:|   |    :|    : ||  :|:|  :  :  :| |
Db        596 PA----LKKADIGIAVDDATDAARSASDIVLTEPGLSVIVSAVLTS--RAIFQRMKNY-- 647

Qy        578 TDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMALKLAMIVGI 632
                              |: |::  | : :   :   :|  | :  : :|: |
Db        648 ------------------TIYAVSITIRIVLGFMLVALIWEFDFSPFMVLIIAI 683
```

Search completed: July 22, 2003, 07:03:00
Job time : 14.0779 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 06:19:03 ; Search time 33.5931 Seconds
(without alignments)
5287.169 Million cell updates/sec

Title: US-09-873-737A-4
Perfect score: 4528
Sequence: 1 MTGRARARARGRARGQETVQ.....VGQSIHREPNLSLSNRLYYL 862

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4515	99.7	862	11	Q9JMB7	Q9jmb7 mus musculu
2	4317.5	95.4	861	4	Q96J94	Q96j94 homo sapien
3	4308.5	95.2	861	4	Q8TBY5	Q8tby5 homo sapien
4	4301.5	95.0	861	4	O95404	O95404 homo sapien
5	3957	87.4	775	4	Q96JD5	Q96jd5 homo sapien
6	3045	67.2	858	13	Q8UVX0	Q8uvx0 brachydanio
7	2017.5	44.6	854	5	Q9GPA8	Q9gpa8 strongyloce
8	1825.5	40.3	808	5	Q9GPA7	Q9gpa7 strongyloce
9	1739.5	38.4	971	11	Q9JMB6	Q9jmb6 mus musculu
10	1725.5	38.1	973	4	Q8TC59	Q8tc59 homo sapien
11	1459.5	32.2	866	5	O76922	O76922 drosophila
12	1398.5	30.9	843	5	Q9VKM1	Q9vkm1 drosophila
13	1360.5	30.0	580	11	Q99MV6	Q99mv6 mus musculu
14	1290	28.5	824	5	P90786	P90786 caenorhabdi
15	1190	26.3	530	4	Q96SW6	Q96sw6 homo sapien
16	1187	26.2	421	5	Q9GPA6	Q9gpa6 strongyloce
17	1137	25.1	722	5	Q17567	Q17567 caenorhabdi
18	932	20.6	371	4	Q9NW28	Q9nw28 homo sapien
19	715	15.8	781	5	Q9U5C9	Q9u5c9 paramecium
20	522.5	11.5	1014	10	Q9SHF3	Q9shf3 arabidopsis
21	520	11.5	1194	10	Q9SHF2	Q9shf2 arabidopsis
22	489.5	10.8	891	5	O16720	O16720 caenorhabdi
23	489.5	10.8	1000	5	Q20578	Q20578 caenorhabdi
24	484	10.7	860	4	Q9H9G7	Q9h9g7 homo sapien
25	481	10.6	1123	10	Q9LP83	Q9lp83 arabidopsis
26	477	10.5	950	5	Q9V3L2	Q9v3l2 drosophila
27	477	10.5	984	5	Q9V6V6	Q9v6v6 drosophila
28	473.5	10.5	924	4	Q9HCK5	Q9hck5 homo sapien
29	470.5	10.4	924	10	Q9ZVD5	Q9zvd5 arabidopsis
30	457	10.1	851	4	Q8TCZ5	Q8tcz5 homo sapien
31	457	10.1	904	10	Q9SDG8	Q9sdg8 oryza sativ
32	449	9.9	990	10	Q9C793	Q9c793 arabidopsis
33	428.5	9.5	1035	5	Q21079	Q21079 caenorhabdi
34	427	9.4	640	4	Q9H1U6	Q9h1u6 homo sapien
35	423	9.3	1037	5	Q9TW94	Q9tw94 caenorhabdi
36	419.5	9.3	922	3	Q96UM0	Q96um0 blumeria gr
37	417.5	9.2	887	10	O48771	O48771 arabidopsis
38	413	9.1	962	5	Q9VUQ6	Q9vuq6 drosophila
39	403.5	8.9	601	5	Q960T1	Q960t1 drosophila
40	394.5	8.7	585	4	Q8WV58	Q8wv58 homo sapien
41	381.5	8.4	898	5	Q9XVI3	Q9xvi3 caenorhabdi
42	374	8.3	694	5	Q8T3N2	Q8t3n2 drosophila
43	363	8.0	377	4	Q96ID1	Q96id1 homo sapien
44	363	8.0	377	11	Q8R3Q7	Q8r3q7 mus musculu
45	356	7.9	938	3	Q9P8T1	Q9p8t1 neurospora

ALIGNMENTS

RESULT 1
Q9JMB7
ID Q9JMB7 PRELIMINARY; PRT; 862 AA.
AC Q9JMB7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE MIWI (piwi).
GN PIWIL1 OR MIWI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Miyagawa S.K., Kimura T., Nakano T.;
RT "Molecular Cloning and characterization of Piwi family genes.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Deng W., Lin H.;
RT "Miwi, a murine homolog of piwi, encodes a cytoplasmic protein
RT essential for spermatogenesis.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB032604; BAA93705.1; -.
DR EMBL; AF438405; AAL31014.1; -.
DR MGD; MGI:1928897; Piwil1.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
SQ SEQUENCE 862 AA; 98574 MW; 45588D13284CCC4C CRC64;

Query Match 99.7%; Score 4515; DB 11; Length 862;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 857; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEGDLVGRGRQRGMVVGAT 60

```
Db          1 MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEGDLVGRGRQRGMVVGAT 60

Qy         61 SKSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHF 120

Db         61 SKSQELQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVKLSTNHF 120

Qy        121 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVT 180

Db        121 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVT 180

Qy        181 EVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPI 240

Db        181 EVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPSDPI 240

Qy        241 DIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQV 300

Db        241 DIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQV 300

Qy        301 SKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQ 360

Db        301 SKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQ 360

Qy        361 PVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 420

Db        361 PVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 420

Qy        421 EVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWS 480

Db        421 EVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWS 480

Qy        481 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTE 540

Db        481 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTE 540

Qy        541 AYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 600

Db        541 AYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 600

Qy        601 IATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWF 660

Db        601 IATKIALQMNCKMGGELWRVDMPLKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWF 660

Qy        661 SRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQ 720

Db        661 SRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQ 720

Qy        721 FLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 780

Db        721 FLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 780

Qy        781 VSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLA 840

Db        781 VSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKLCHVYYNWPGVIRVPAPCQYAHKLA 840

Qy        841 FLVGQSIHREPNLSLSNRLYYL 862

Db        841 FLVGQSIHREPNLSLSNRLYYL 862
```

```
RESULT 2
Q96J94
ID   Q96J94      PRELIMINARY;      PRT;   861 AA.
AC   Q96J94;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   PIWI protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Sha J.H.;
RT   "Cloning and identification of human piwi protein related to testis
RT   development.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF387507; AAK69348.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   861 AA;  98603 MW;  58D7F6C7321DEFA4 CRC64;
```

Query Match 95.4%; Score 4317.5; DB 4; Length 861;
Best Local Similarity 94.3%; Pred. No. 4.9e-316;
Matches 813; Conservative 22; Mismatches 26; Indels 1; Gaps 1;

```
Qy          1 MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEGDLVGRGRQRGMVVGAT 60

Db          1 MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRG-TAGGT 59

Qy         61 SKSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHF 120

Db         60 AKSQGLQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHF 119

Qy        121 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVT 180

Db        120 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVT 179

Qy        181 EVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPI 240

Db        180 EVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPI 239

Qy        241 DIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQV 300

Db        240 DIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQV 299

Qy        301 SKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQ 360

Db        300 SKELIGLVVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQ 359

Qy        361 PVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 420

Db        360 PVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 419

Qy        421 EVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWS 480

Db        420 EVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWS 479

Qy        481 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTE 540

Db        480 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDRTE 539

Qy        541 AYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 600

Db        540 AYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 599

Qy        601 IATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWF 660

Db        600 IATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWF 659

Qy        661 SRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQ 720

Db        660 SRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQ 719

Qy        721 FLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 780

Db        720 FLDCLKSIGRGYNPRLTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 779

Qy        781 VSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLA 840

Db        780 VSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLA 839
```

```
Qy        841 FLVGQSIHREPNLSLSNRLYYL 862
              |||||||||||||||||||||||
Db        840 FLVGQSIHREPNLSLSNRLYYL 861


RESULT 3
Q8TBY5
ID   Q8TBY5      PRELIMINARY;      PRT;   861 AA.
AC   Q8TBY5;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Piwi-like 1 (Drosophila).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Strausberg R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC028581; AAH28581.1; -.
SQ   SEQUENCE   861 AA;  98545 MW;  D33376EEDEDF743A CRC64;

  Query Match             95.2%;  Score 4308.5;  DB 4;  Length 861;
  Best Local Similarity   94.2%;  Pred. No. 2.3e-315;
  Matches  812;  Conservative   21;  Mismatches   28;  Indels    1;  Gaps    1;

Qy          1 MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEGDLVGRGRQRGMVVGAT 60
              |||||||||||||||||| | ||: |||||||| ||||  | ||:| |||||||   | |
Db          1 MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRG-TAGGT 59

Qy         61 SKSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHF 120
              :||| |||||||||||||||||||||| |||||||||||||||||||||:||||||
Db         60 AKSQGLQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHF 119

Qy        121 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVT 180
              |||||||||||||||||||||||||||||||||||||:||||||||||||||| |||
Db        120 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVT 179

Qy        181 EVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPI 240
              ||||:||||| |||||||||||||||||||||| |||||||||||||||||||:|||
Db        180 EVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPI 239

Qy        241 DIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQV 300
              |||:|||||||||||||||||:|||||||||||||||||||| | ||||||||||
Db        240 DIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQV 299

Qy        301 SKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQ 360
              |||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 SKELIGLVVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQ 359

Qy        361 PVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 420
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        360 PVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 419

Qy        421 EVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWS 480
              ||||||||||:|||||||||||||||||||||||||:||||||||||||||||||||
Db        420 EVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWS 479

Qy        481 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTE 540
              |||||||||| ||||||||||||||||||||||||||||:|||||||||||||||||
Db        480 KETRGAPLISVNPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMKKAIMIEVDDRTE 539

Qy        541 AYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 600
              |||| ||||||:|||||||||||||||||||| ||||||||||||||||||||||||||
Db        540 AYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYPCTDCPTPSQCVVARTLGKQQTVMA 599

Qy        601 IATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWF 660
              |||||||||||||||||||: ||| |||||||||| ||||||||||||||||||||||||
Db        600 IATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWF 659

Qy        661 SRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQ 720
              |||:|||||||||||||||||||||||: |||||||:|||||||||||||||||||||||
Db        660 SRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQ 719

Qy        721 FLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 780
              |||||||:|||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        720 FLDCLKSIGRGYNPRLTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 779

Qy        781 VSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLA 840
              ||||||||||||||||||||:||||||||||||| ||:|||||||||||||||||||||
Db        780 VSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLA 839

Qy        841 FLVGQSIHREPNLSLSNRLYYL 862
              |||||||||||||||||||||||
Db        840 FLVGQSIHREPNLSLSNRLYYL 861


RESULT 4
O95404
ID   O95404      PRELIMINARY;      PRT;   861 AA.
AC   O95404;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   HIWI.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE=99069219; PubMed=9851978;
RA   Cox D.N., Chao A., Baker J., Chang L., Qiao D., Lin H.;
RT   "A novel class of evolutionarily conserved genes defined by piwi are
RT   essential for stem cell self-renewal.";
RL   Genes Dev. 12:3715-3727(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Qiao D., Zeeman A.-M., Deng W., Looijenga L.H.J., Lin H.;
RT   "Molecular characterization of hiwi, a human member of the piwi stem
RT   cell gene family whose overexpression is correlated to seminomas.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF104260; AAC97371.2; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   861 AA;  98530 MW;  43D7F60E99D997B7 CRC64;

  Query Match             95.0%;  Score 4301.5;  DB 4;  Length 861;
  Best Local Similarity   94.1%;  Pred. No. 7.8e-315;
  Matches  811;  Conservative   22;  Mismatches   28;  Indels    1;  Gaps    1;

Qy          1 MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEGDLVGRGRQRGMVVGAT 60
              |||||||||||||||||| | ||: |||||||| ||||  | ||:| |||||||   | |
Db          1 MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRG-TAGGT 59

Qy         61 SKSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHF 120
              :||| |||||||||||||||||||||| |||||||||||||||||||||:||||||
Db         60 AKSQGLQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHF 119

Qy        121 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVT 180
              |||||||||||||||||||||||||||||||||||||:||||||||||||||| |||
Db        120 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVT 179

Qy        181 EVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPI 240
              ||||:||||| |||||||||||||||||||||| |||||||||||||||||||:|||
Db        180 EVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPI 239

Qy        241 DIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQV 300
              |||:|||||||||||||||||:|||||||||||||||||||| | ||||||||||
```

```
Db              ALVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQV 299

              ELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQ 360
                     :      |||||||||||||||||||||||| ||||||||||||||||||
            J SKELIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQ 359

Q.        361 PVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 420
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        360 PVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 419

Qy        421 EVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWS 480
              ||||||||||:|||||||||||||||||||||||||||||:||:||||||||||||||
Db        420 EVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWS 479

Qy        481 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTE 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDRTE 539

Qy        541 AYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 600
              |||| |||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 AYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 599

Qy        601 IATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWF 660
              ||||||||||||||||||: ||| |||||||||| ||||||||||||||||||||||||
Db        600 IATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWF 659

Qy        661 SRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQ 720
              |||:|||||||||||||||||||||||: ||||||||:||||||||||||||||||||||
Db        660 SRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQ 719

Qy        721 FLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 780
              |||||||:|||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        720 FLDCLKSIGRGYNPRLTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 779

Qy        781 VSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLA 840
              ||||||||||||||||||:|||||||||||||| ||:|||||||||||||||||||||||
Db        780 VSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLA 839

Qy        841 FLVGQSIHREPNLSLSNRLYYL 862
              ||||||||||||||||||||||
Db        840 FLVGQSIHREPNLSLSNRLYYL 861
```

RESULT 5
Q96JD5
ID Q96JD5 PRELIMINARY; PRT; 775 AA.
AC Q96JD5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE HIWI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Sharma A.K., Nelson M.C., Brandt J.E., Wessman M., Muhmud N.,
RA Weller K.P., Hoffman R.;
RT "Human CD34+ Stem Cells Express the hiwi Gene, a Human Homolog of the
RT Drosophila Gene piwi.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF264004; AAK92281.1; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
SQ SEQUENCE 775 AA; 89484 MW; DF169A2E9EAFD916 CRC64;

Query Match 87.4%; Score 3957; DB 4; Length 775;
Best Local Similarity 95.9%; Pred. No. 5.6e-289;
Matches 740; Conservative 19; Mismatches 13; Indels 0; Gaps 0;

```
Qy         91 GVNTRQNLDHVKESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSAL 150
              |||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db          4 GVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSAL 63

Qy        151 LFQHEDLIGRCHAFDGTILFLPKRLQHKVTEVFSQTRNGEHVRITITLTNELPPTSPTCL 210
              |||||||||:|||||||||||||||| |||||||:|||| ||||||||||||||||||||
Db         64 LFQHEDLIGKCHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCL 123

Qy        211 QFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDV 270
              ||||| |||||||||||||||||:||||||:||||||||||||||||||:|||||||||
Db        124 QFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDV 183

Qy        271 SHKVLRSETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPK 330
              |||||||||||||| | ||||||||||||||||:||||||||||||||||||||||||||
Db        184 SHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKELIGLVVLTKYNNKTYRVDDIDWDQNPK 243

Qy        331 STFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMXIPELCYL 390
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        244 STFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYL 303

Qy        391 TGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNL 450
              |||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db        304 TGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNL 363

Qy        451 LSFSGRILQSEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAAN 510
              |||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 LSFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAAN 423

Qy        511 SLIQNLFKVTPAMGIQMKKAIMIEVDDRTEAYLRALQQKVTSDTQIVVCLLSSNRKDKYD 570
              |||||||||||||||:||:|||||||||||||| |||||||:|||||||||||||||||
Db        424 SLIQNLFKVTPAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYD 483

Qy        571 AIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMALKLAMIV 630
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 AIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIV 543

Qy        631 GIDCYHDTTAGRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGC 690
              ||||||| |||||||||||||||||||:|||||||||||||||||||||||||||||: |
Db        544 GIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSC 603

Qy        691 NEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFF 750
              |||||||:|||||||||||||||||||||||||||:|||||||||||||||||| |||
Db        604 NEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVVKKRVNTRFF 663

Qy        751 AQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQ 810
              ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db        664 AQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQ 723

Qy        811 RLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 862
              ||||| ||:|||||||||||||||||||||||||||||||||||||||||||
Db        724 RLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 775
```

RESULT 6
Q8UVX0
ID Q8UVX0 PRELIMINARY; PRT; 858 AA.
AC Q8UVX0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Piwi protein.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Weeraratne S.D., Gong Z., Tan C.-H.;
RT "Cloning and characterization of zebrafish homolog of piwi, essential
RT for germ-line stem cell self-renewal.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.

```
DR   EMBL; AF336369; AAL57170.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   858 AA;  97451 MW;  6A12F2E511465777 CRC64;

  Query Match             67.2%;  Score 3045;  DB 13;  Length 858;
  Best Local Similarity   64.4%;  Pred. No. 2.8e-220;
  Matches  560;  Conservative  131;  Mismatches  160;  Indels   18;  Gaps    4;

Qy          1 MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQ-------SPTEGDLVGRGRQR 53
              |||||||||:||| ||||          ||  ||  |:       :|:|| |||||||:
Db          1 MTGRARARSRGRGRGQEPA---------APGAQPPVSQEAAKPVVSTPSEGQLVGRGRQK 51

Qy         54 GMVVGATSKSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIV 113
                 || |:   |||||||:: : |||||||||| ||:||| ::|||||||| ||   :
Db         52 -PAPGAMSEEAMLQISAGFQQVKIGERGGRRRDFHDSGVHTRQLMEHVKESKTGVSGTAI 110

Qy        114 KLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPK 173
              :|  |  || ||| |||||||:|| | ||:||||||||||||: :|: | ||| |||||
Db        111 ELRANFMRLLSRPMWALYQYHVDYKPPMESRRLRSALLFQHEETLGKAHTFDGAILFLPN 170

Qy        174 RLQHKVTEVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNY 233
              :|::  | : |:|||| | ||:||||||||:|| ||||||| |||:|:|:|:|||||:|
Db        171 KLRNAETVLCSETRNGEKVEITVTLTNELPPSSPVCLQFYNILFRRILRILNMQQIGRHY 230

Qy        234 YNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEE 293
              ||| || :|| ||| ||||| |:|||||::||||:|||||||||||||||::| ||  :
Db        231 YNPDDPFNIPQHRLTIWPGFMTTILQYESSIMLCSDVSHKVLRSETVLDFMYSLRQQCGD 290

Qy        294 HKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQ 353
              :| |  :|||:|||:|||||||||||:||| ||  | :|||| | :|:|| |:: ||
Db        291 QRFPEACTKELVGLIILTKYNNKTYRIDDIAWDHTPNNTFKKGD-TEISFKNYFKSQYGL 349

Qy        354 EITDLKQPVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRL 413
              :|||  | :|||  ||    |  ||||| :|| ||||||||||| |||:||||| ||||
Db        350 DITDGNQVLLVSHVKRLGPSGRPPPGPAMLVPEFCYLTGLTDKMRADFNIMKDLASHTRL 409

Qy        414 TPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYN 473
              :||||:  : |||  |::: :|| ||  ||||:: ||| :||:| ||:| |||: |:||
Db        410 SPEQREGRINRLISNINRNGDVQNELTTWGLSFENKLLSLNGRVLPSERIIQGGRAFEYN 469

Qy        474 PQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMI 533
              |  |||||| || ||||   |||||: ||||| : | ||:| | ||:  |||:|:::|:||
Db        470 PWTADWSKEMRGLPLISCMSLDNWLMFYTRRNADVAQSLLQTLNKVSGPMGIRMQRAVMI 529

Qy        534 EVDDRTEAYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLG 593
              | :|| |: |||||| |  :||:|| :| :|||||| :||||| |||||||||:||:
Db        530 EYEDRQESLLRALQQNVARETQMVVVILPTNRKDKYDCVKKYLCVDCPTPSQCVVSRTIS 589

Qy        594 KQQTVMAIATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASIN 653
              | | :| :|||||||||||||||||| |:: |:  |||||||||| ||:||   |||:|
Db        590 KPQALMTVATKIALQMNCKMGGELWSVEIPLRQLMIVGIDCYHDTAAGKRSIGAMVASLN 649

Qy        654 EGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTL 713
              :||:||||:|| |:||||::| ||  || ||:|:   |  :|||:|||||||| |:::
Db        650 QGMSRWFSKCVLQNRGQEIIDALKGSLQGALKAYLKYNNSLPSRIIVYRDGVGDGMLQSV 709

Qy        714 VNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRP 773
              |:||||| :  :|::|: | |:|:|:|||||:::|||||:  |:: || |||||| |||||
Db        710 VDYEVPQIMQSIKTMGQDYEPKLSVVVVKKRISSRFFARIDGKIANPPPGTVIDTEVTRP 769

Qy        774 EWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPC 833
              |||||||||||| | |:|||||::|:|||||||:|||||| ||:|||| |::|||||||
Db        770 EWYDFFIVSQAVRFGCVAPTHYNVVFDNSGLKPDHMQRLTYKLCHMYYNWQGIVRVPAPC 829

Qy        834 QYAHKLAFLVGQSIHREPNLSLSNRLYYL 862
              |||||||||||||||:|||::| : ||||
Db        830 QYAHKLAFLVGQSIHKEPNMNLDDFLYYL 858

RESULT 7
```

```
Q9GPA8
ID   Q9GPA8      PRELIMINARY;      PRT;   854 AA.
AC   Q9GPA8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Seawi.
OS   Strongylocentrotus purpuratus (Purple sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Strongylocentrotus.
OX   NCBI_TaxID=7668;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rodriguez A.J., Bonder E.M.;
RT   "Seawi - Cloning and Molecular Characterization of a Sea Urchin
RT   Homolog of Piwi.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY014899; AAG42533.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   854 AA;  96722 MW;  509A1D39C0D1922C CRC64;

  Query Match             44.6%;  Score 2017.5;  DB 5;  Length 854;
  Best Local Similarity   46.9%;  Pred. No. 5.7e-143;
  Matches  403;  Conservative  152;  Mismatches  272;  Indels   33;  Gaps   12;

Qy         28 QQPGYIPPRPQQS---PTEGD----LVGRGRQRGMVV----GATSKSQEL-QISAGFQEL 75
              ::||   | | |    | |      : |||| ||       || | :  ::::   :|
Db          3 RRPGEQAPPPMQQAAPPQEASSGMPVPGRGRSRGGTTQPGPGAVPKQEPRGELTSQIAKL 62

Qy         76 SLAERG----------GRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHFRLTS 124
              ::   |           ||||  :   ||   : ||::  |     : |  | |:| :
Db         63 AIGASGEGRVGGAVQDRRRRRDEIELEPTTRPE-NFVKQAIAGDK---IALIANGFKLKT 118

Qy        125 RPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVTEVFS 184
              :| | |||| :|: | :   | | |||  |  |:|:    |   ::   :|  ||| : :
Db        119 KPDWQLYQYRVDFEPEILNPRARFALLKGHSALLGKGLTLDMDTMYSLFKLAEKVTNLSA 178

Qy        185 QTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIPN 244
              : ::| :| | :|    | | :|  |  ||: ||| ||::|:::|:|||||:|:  |||
Db        179 ERKDGSNVNICVTHVATLNPMAPNTLHLYNVLFRRCLKMINMEQVGRNYYDPTAAIDIKQ 238

Qy        245 HRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQVSKEL 304
              | | :|||| ||||||| ::|| :|:|||||||::|| : | :|: :    :|:|:::| :
Db        239 HGLQLWPGFVTSILQYEYDVMLLSDISHKVLRTQTVWEVMNDLFNKA-RGRFKEEITKLM 297

Qy        305 IGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLV 364
              || |||||||||||||||||:|  |  ||:   |  ||:::|::|  | : | |: ||:||
Db        298 IGQIVLTKYNNKTYRVDDIDFDTTPADTFETRSG-PVSYVDYFKKSYERVIHDVNQPMLV 356

Qy        365 SQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGR 424
              |:||:|    |   |||   ||||:||||:| || |||||||:  ||| |: | | :
Db        357 SRPKKREEKKGV--GPAYLPPELCFLTGLSDDMRADFNVMKDMGQHTRVGPQDRCRTLSG 414

Qy        425 LIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWSKETR 484
              |  :  :: |:  |  ||: ||   :  :||:|  ||: | || | |||  ||||::||
Db        415 FIKKLSSNEEVKTYLDSWGMEFDKEQVKLTGRVLPPEKLFQRGKQFSYNPSNADWSRDTR 474

Qy        485 GAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEV-DDRTEAYL 543
              |  |   | |:|| : ||||:       |::| :|   ||: ::   ::|: ||||| |
Db        475 GNALTDAKILNNWKIFYTRRDANRGQDFIKSLVRVANPMGMNVRGPEIVELPDDRTETYT 534

Qy        544 RALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIAT 603
              |:|| ::  |||||| :| :||||:|||||      | ||| :|:||| ||| :|::||
Db        535 RSLQAQIAQDTQIVVVILPTNRKDRYDAIKKTCVVTHPCPSQVIVSRTLSKQQMLMSVAT 594

Qy        604 KIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWFSRC 663
              |||:|||||||:||||:: |   ||:||| |||:    ||: |||||:|:  | :|| |
Db        595 KIAMQMNCKMGGDLWRVEIPLSNLMIIGIDSYHDSLTKGRSVLGFVASMNKSQTSFFSSC 654
```

```
Qy        664 VFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLD 723
              ||     |    |   :  ||: :   ||  | |:|::|||||| |:  :|:||: |  |
Db        655 AFQHAQGEFGANLSTLMNNALKRYYQINEKFPERIIIFRDGVGDSQVNLVVDYELKQIKD 714

Qy        724 CLKSV-GRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVS 782
              |  |  :|   :| |:|||||:| ||||   | | || ||||||  ||:|  ||:||:|
Db        715 TLDKVYPQGTVHKLAVVVVKKRINNRFFANLRGGLSNPPPGTVIDDVVTKPHLYDYFIIS 774

Qy        783 QAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFL 842
              |:|| |||||| |||:||::|||||:|||||  |:|:|||| :|||||| ||||||||
Db        775 QSVRQGSVSPTSYNVVYDTTGLKPDHMQRLTYKLTHLYFNWPGTVRVPAPCMYAHKLAFL 834

Qy        843 VGQSIHREPNLSLSNRLYYL 862
              :|||:|  ||  |:: |:||
Db        835 IGQSVHEVPNNRLADTLFYL 854


RESULT 8
Q9GPA7
ID   Q9GPA7      PRELIMINARY;      PRT;   808 AA.
AC   Q9GPA7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Seawi (Fragment).
OS   Strongylocentrotus purpuratus (Purple sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Strongylocentrotus.
OX   NCBI_TaxID=7668;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rodriguez A.J., Bonder E.M.;
RT   "Seawi - Cloning and Molecular Characterization of a Sea Urchin
RT   Homolog of Piwi.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY014900; AAG42534.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
FT   NON_TER     808    808
SQ   SEQUENCE   808 AA;  91366 MW;  19528F4B9D10474C CRC64;

  Query Match             40.3%;  Score 1825.5;  DB 5;  Length 808;
  Best Local Similarity   45.8%;  Pred. No. 1.5e-128;
  Matches  372;  Conservative  144;  Mismatches  263;  Indels   33;  Gaps   12;

Qy         28 QQPGYIPPRPQQS---PTEGD----LVGRGRQRGMVV----GATSKSQEL-QISAGFQEL 75
              ::||   | | |    | |      : |||| ||       ||  | :   ::::   :|
Db          3 RRPGEQAPPPMQQAAPPQEASSGMPVPGRGRSRGGTTQPGPGAVPKQEPRGELTSQIAKL 62

Qy         76 SLAERG----------GRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHFRLTS 124
              ::   |           ||||  :    ||   : ||::  |     : |  | |:| :
Db         63 AIGASGEGRVGGAVQDRRRRRDEIELEPTTRPE-NFVKQAIAGDK---IALIANGFKLKT 118

Qy        125 RPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVTEVFS 184
              :| | |||| :|: | :   | | |||  |  |:|:   |   ::   :|  ||| : :
Db        119 KPDWQLYQYRVDFEPEILNPRARFALLKGHSALLGKGLTLDMDTMYSLFKLAEKVTNLSA 178

Qy        185 QTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIPN 244
              : ::| :| | :|    | | :|  |  ||: ||| ||::|::|:|||||:|:  |||
Db        179 ERKDGSNVNICVTHVATLNPMAPNTLHLYNVLFRRCLKMINMEQVGRNYYDPTAAIDIKQ 238

Qy        245 HRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQVSKEL 304
              | | :|||| ||||||| ::|| :|:|||||||::|| : | :|: :    :|:|:::| :
Db        239 HGLQLWPGFVTSILQYEYDVMLLSDISHKVLRTQTVWEVMNDLFNKA-RGRFKEEITKLM 297

Qy        305 IGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLV 364
              || |||||||||||||||||:|  | ||:  |  ||:::|::| | : | |: ||:||
Db        298 IGQIVLTKYNNKTYRVDDIDFDTTPADTFETRSG-PVSYVDYFKKSYERVIHDVNQPMLV 356

Qy        365 SQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGR 424
              |:||:|    |  |||  ||||:||||:| || |||||||:  |||: |: | | :
Db        357 SRPKKREEKKGV--GPAYLPPELCFLTGLSDDMRADFNVMKDMGQHTRVGPQDRCRTLSG 414

Qy        425 LIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWSKETR 484
              |  :  :: |:  |  ||: ||   :  :||:|  ||: | || | |||  ||||::||
Db        415 FIKKLSSNEEVKTYLDSWGMEFDKEQVKLTGRVLPPEKLFQRGKQFSYNPSNADWSRDTR 474

Qy        485 GAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEV-DDRTEAYL 543
              |  |   | |:|| : ||||:       |::| :|   ||: ::   ::|: ||||| |
Db        475 GNALTDAKILNNWKIFYTRRDANRGQDFIKSLVRVANPMGMNVRGPEIVELPDDRTETYT 534

Qy        544 RALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIAT 603
              |:|| ::  |||||| :| :||||:||||||       | ||| :|:||| ||| :|::||
Db        535 RSLQAQIAQDTQIVVVILPTNRKDRYDAIKKTCVVTHPCPSQVIVSRTLSKQQMLMSVAT 594

Qy        604 KIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWFSRC 663
              |||:|||||||:||||:: |   ||:||| |||:    ||: |||||:|:  | :|| |
Db        595 KIAMQMNCKMGGDLWRVEIPLSNLMIIGIDSYHDSLTKGRSVLGFVASMNKSQTSFFSSC 654

Qy        664 VFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLD 723
              ||     |    |   :  ||: :   ||  | |:|::|||||| |:  :|:||: |  |
Db        655 AFQHAQGEFGANLSTLMNNALKRYYQINEKFPERIIIFRDGVGDSQVNLVVDYELKQIKD 714

Qy        724 CLKSV-GRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVS 782
              |  |  :|   :| |:|||||:| ||||   | | || ||||||  ||:|  ||:||:|
Db        715 TLDKVYPQGTVHKLAVVVVKKRINNRFFANLRGGLSNPPPGTVIDDVVTKPHLYDYFIIS 774

Qy        783 QAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTY 814
              |:|| |||||| |||:||::|||||:|||||
Db        775 QSVRQGSVSPTSYNVVYDTTGLKPDHMQRLTY 806


RESULT 9
Q9JMB6
ID   Q9JMB6      PRELIMINARY;      PRT;   971 AA.
AC   Q9JMB6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   MILI (Miwi like).
GN   PIWIL2 OR PIWIL1L OR MILI.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Miyagawa S.K., Kimura T., Nakano T.;
RT   "Molecular Cloning and characterization of Piwi family genes.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB032605; BAA93706.1; -.
DR   MGD; MGI:1930036; Piwil2.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   971 AA;  109488 MW;  01E143C6513310FB CRC64;

  Query Match             38.4%;  Score 1739.5;  DB 11;  Length 971;
  Best Local Similarity   38.8%;  Pred. No. 5.8e-122;
  Matches  341;  Conservative  176;  Mismatches  293;  Indels   69;  Gaps   14;

Qy          7 ARARGRARGQETVQHVGAAAS----------------QQPGYIPPRPQQSPTEGDLVGR 49
              :|  ||   : ::  :| |||                 : |  :|  |  |||
Db        139 SRMLGRGSSEVSLLPLGRAASSIGRGMDKPPSAFGLTARDPPRLPQPPALSPTS----LH 194

Qy         50 GRQRGMVVGATSKSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSS 109
                    |:    | :|| :                                  | ||
Db        195 SADPPPVLTMERKEKELLV----------------------------------KQGSK 218

Qy        110 GIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTIL 169
              |    |  |  ::    : |:|||||: ::| :| : :|  :|  |: : |   ||||:||
Db        219 GTPQSLGLNLIKIQCHNE-AVYQYHVTFSPSVECKSMRFGMLKDHQSVTGNVTAFDGSIL 277
```

```
Qy        170 FLPKRLQHKVTEVFSQTRNGE-HVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQ 228
Db        278 YLPVKLQ-QVVELKSQRKTDDAEISIKIQLTKILEPCSDLCIPFYNVVFRRVMKLLDMKL 336

Qy        229 IGRNYYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLY 288
Db        337 VGRNFYDPTSAMVLQQHRLQIWPGYAASIRRTDGGLFLLADVSHKVIRNDSVLDVMHAIY 396

Qy        289 QQTEEHKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYR 348
Db        397 QQNKEH-FQDECSKLLVGSIVITRYNNRTYRIDDVDWNKTPKDSFVMSDGKEITFLEYYS 455

Qy        349 KQYNQEITDLKQPVLVSQP-KRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDL 407
Db        456 KNYGITVKEDDQPLLIHRPSERQNNHGMLLKGEILLLPELSFMTGIPEKMKKDFRAMKDL 515

Qy        408 AVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGG 467
Db        516 TQQINLSPKQHHGALECLLQRISQNETASNELTRWGLSLHKDVHKIEGRLLPMERINLRN 575

Qy        468 KTFDYNPQFADWSKE-TRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQ 526
Db        576 TSF-VTSEDLNWVKEVTRDASILTI-PMHFWALFYPKRAMDQARELVNMLEKIAGPIGMR 633

Qy        527 MKKAIMIEV-DDRTEAYLRALQQ--KVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTP 583
Db        634 ISPPAWVELKDDRIETYIRTIQSLLGVEGKIQMVVCIIMGTRDDLYGAIKKLCCVQSPVP 693

Qy        584 SQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRR 643
Db        694 SQVINVRTIGQPTRLRSVAQKILLQMNCKLGGELWGVDIPLKQLMVIGMDVYHDPSRGMR 753

Qy        644 SIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRD 703
Db        754 SVVGFVASINLTLTKWYSRVVFQMPHQEIVDSLKLCLVGSLKKYYEVNHCLPEKIVVYRD 813

Qy        704 GVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLPG 763
Db        814 GVSDGQLKTVANYEIPQLQKCFEAFD-NYHPKMVVFVVQKKISTNLYLAAPDHFVTPSPG 872

Qy        764 TVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNW 823
Db        873 TVVDHTITSCEWVDFYLLAHHVRQGCGIPTHYICVLNTANLSPDHMQRLTFKLCHMYWNW 932

Qy        824 PGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 862
Db        933 PGTIRVPAPCKYAHKLAFLSGQILHHEPAIQLCGNLFFL 971
```

RESULT 10
Q8TC59
ID Q8TC59 PRELIMINARY; PRT; 973 AA.
AC Q8TC59;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to piwi like homolog 1 (Drosophila)-like.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC025995; AAH25995.1; -.
SQ SEQUENCE 973 AA; 109848 MW; C44398136B144CA0 CRC64;

Query Match 38.1%; Score 1725.5; DB 4; Length 973;
Best Local Similarity 38.6%; Pred. No. 6.5e-121;
Matches 340; Conservative 171; Mismatches 298; Indels 71; Gaps 13;

```
Qy          7 ARARGRARGQETVQHVGAAA----------------SQQPGYI--PPRPQQSPTEGDLV 47
Db        141 SRTLGRGSSDASLLPLGRAAGGISREVDKPPCTFSTPSRGPPQLSSPPALPQSPLHSP-- 198

Qy         48 GRGRQRGMVVGATSKSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTG 107
Db        199 ----DRPLVLTVEHKEKELIV-----------------------------------KQG 218

Qy        108 SSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGT 167
Db        219 SKGTPQSLGLNLVKIQCHNE-AVYQYHVTFSPNVECKSMRFGMLKDHQAVTGNVTAFDGS 277

Qy        168 ILFLPKRLQHKVTEVFSQTR-NGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNL 226
Db        278 ILYLPVKLQ-QVLELKSQRKTDSAEISIKIQMTKILEPCSDLCIPFYNVVFRRVMKLLDM 336

Qy        227 QQIGRNYYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFN 286
Db        337 KLVGRNFYDPTSAMVLQQHRLQIWPGYAASIRRTDGGLFLLADVSHKVIRNDCVLDVMHA 396

Qy        287 LYQQTEEHKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEY 346
Db        397 IYQQNKEH-FQDECTKLLVGNIVITRYNNRTYRIDDVDWNKTPKDSFTMSDGKEITFLEY 455

Qy        347 YRKQYNQEITDLKQPVLVSQPKRRR-GPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMK 405
Db        456 YSKNYGITVKEEDQPLLIHRPSERQDNHGMLLKGEILLLPELSFMTGIPEKMKKDFRAMK 515

Qy        406 DLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQ 465
Db        516 DLAQQINLSPKQHHSALECLLQRIAKNEAATNELMRWGLRLQKDVHKIEGRVLPMERINL 575

Qy        466 GGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGI 525
Db        576 KNTSF-ITSQELNWVKEVTRDPSILTIPMHFWALFYPKRAMDQARELVNMLEKIAGPIGM 634

Qy        526 QMKKAIMIEV-DDRTEAYLRALQQKVTSD--TQIVVCLLSSNRKDKYDAIKKYLCTDCPT 582
Db        635 RMSPPAWVELKDDRIETYVRTIQSTLGAEGKIQMVVCIIMGPRDDLYGAIKKLCCVQSPV 694

Qy        583 PSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGR 642
Db        695 PSQVVNVRTIGQPTRLRSVAQKILLQINCKLGGELWGVDIPLKQLMVIGMDVYHDPSRGM 754

Qy        643 RSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYR 702
Db        755 RSVVGFVASINLTLTKWYSRVVFQMPHQEIVDSLKLCLVGSLKKFYEVNHCLPEKIVVYR 814

Qy        703 DGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLP 762
Db        815 DGVSDGQLKTVANYEIPQLQKCFEAF-ENYQPKMVVFVVQKKISTNLYLAAPQNFVTPTP 873

Qy        763 GTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYN 822
Db        874 GTVVDHTITSCEWVDFYLLAHHVRQGCGIPTHYVCVLNTANLSPDHMQRLTFKLCHMYWN 933

Qy        823 WPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 862
Db        934 WPGTIRVPAPCKYAHKLAFLSGHILHHEPAIQLCENLFFL 973
```

RESULT 11
O76922
ID O76922 PRELIMINARY; PRT; 866 AA.
AC O76922;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE AUB protein.
GN AUB OR STING OR CG6137.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

```
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON S;
RX   MEDLINE=96262192; PubMed=8676882;
RA   Schmidt A., Hollmann M., Schaefer U.;
RT   "A newly identified Minute locus, M(2)32D, encodes the ribosomal
RT   protein L9 in Drosophila melanogaster.";
RL   Mol. Gen. Genet. 251:381-387(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON S;
RA   Schmidt A., Palumbo G., Bozzetti M.P., Tritto P., Pimpinelli S.,
RA   Schaefer U.;
RT   "Genetic and molecular characterization of sting, a gene involved in
RT   crystal formation and meiotic drive in Drosophila melanogaster.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Tsang G., Brokstein P., Frise E., Harvey D., Evans-Holm M.,
RA   Lewis S.E., Suh C., Rubin G.M.;
RT   "Full Length Drosophila melanogaster cDNA sequence.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003630; AAF53046.1; -.
DR   EMBL; X94613; CAA64320.1; -.
DR   EMBL; AF145680; AAD38655.1; -.
DR   FlyBase; FBgn0000146; aub.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   866 AA;  98558 MW;  8F84CF1B0B89103B CRC64;

  Query Match             32.2%;  Score 1459.5;  DB 5;  Length 866;
  Best Local Similarity   36.4%;  Pred. No. 5.7e-101;
  Matches  320;  Conservative  174;  Mismatches  339;  Indels   47;  Gaps   18;

Qy          7 ARARGRARGQETVQHVGAAASQQPGYIPPRPQQS-PTEGDLVGRGRQRGMVVGATSKSQE 65
              || ||| |    |:          |: |   |:| |:  :    | |      |      :
Db         10 ARGRGRGRKPNNVE-------ANRGFAPSLGQKSDPSHSE----GNQAS---GGNGGGGD 55

Qy         66 LQISAGFQELSLAERGGRRRDFHDXG-VNTRQNLDHV-------KESKTGSSGIIVKLST 117
               |:    :: ||:     : :    | |  |: :  :          || |  |  : :
Db         56 AQVGPSIEKSSLSAVQMHKSEGDPRGSVRGRRLITDLVYSRPPGMTSKKGVVGTHITVQA 115

Qy        118 NHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLP---KR 174
              |:|::  || | :||| :|: | :|| ||| : |::|: ::|  : |||| :|     |
Db        116 NYFKVLKRPNWTIYQYRVDFTPDVEATRLRRSFLYEHKGILGG-YIFDGTNMFCINQFKA 174

Qy        175 LQHK--VTEVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRN 232
              :|    | |: :::| ||:: | |     :  |     |  |:  || :: ::|: : |
Db        175 VQDSPYVLELVTKSRAGENIEIKIKAVGSVQSTDAEQFQVLNLILRRAMEGLDLKLVSRY 234

Qy        233 YYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTE 292
              ||:|   |:: | |: :|||: ||| |:||:|:||::: |||:|:||: : : :  : ::
Db        235 YYDPQAKINLENFRMQLWPGYQTSIRQHENDILLCSEICHKVMRTETLYNILSDAIRDSD 294

Qy        293 EHKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYN 352
              :  :|    : ::|:::|| |||||||:||:|:   |   ||  || |:|:::||:|:||
Db        295 D--YQSTFKRAVMGMVILTDYNNKTYRIDDVDFQSTPLCKFKTNDG-EISYVDYYKKRYN 351

Qy        353 QEITDLKQPVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTR 412
              | |||||:::|:|  :   ||     | ||||   ||:|| || ||  :: :: |||
Db        352 IIIRDLKQPLVMSRPTDKNIRGGN-DQAIMIIPELARATGMTDAMRADFRTLRAMSEHTR 410

Qy        413 LTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDY 472
              | |:::|   :      :          |: | :  || |:    |:|  |||   | :
Db        411 LNPDRRIERLRMFNKRLKSCKQSVETLKSWNIELDSALVEIPARVLPPEKILFGNQKIFV 470

Qy        473 NPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIM 532
                 |||: | |   :     :: | :|   ||       :|   :   :| : :   |
Db        471 CDARADWTNEFRTCSMFKNVHINRWYVITPSRNLRETQEFVQMCIRTASSMKMNICNPIY 530

Qy        533 IEV-DDRTEAYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVART 591
              |: |||   | :|:     :| |||: :: |  ::||  |||  | | | ||| |  :
Db        531 EEIPDDRNGTYSQAIDNAAANDPQIVMVVMRSPNEEKYSCIKKRTCVDRPVPSQVVTLKV 590

Qy        592 LGKQQT----VMAIATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAG 647
              :  :|     :|:|||| :||| |: |  |:| : |   | || |  |      :|
Db        591 IAPRQQKPTGLMSIATKVVIQMNAKLMGAPWQVVIPLHGLMTVGFDVCHSPKNKNKSYGA 650

Qy        648 FVASINEGMT-RWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVG 706
              ||::::  : |:||      :|||| : : | :  |||::   :  :| |:: :|||||
Db        651 FVATMDQKESFRYFSTVNEHIKGQELSEQMSVNMACALRSYQEQHRSLPERILFFRDGVG 710

Qy        707 DGQLKTLVNYEVPQFLDCL----KSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLP 762
              ||||  :|| ||    | |    || |:    |:| |:| ||:|:|:|  :| |  ||:|
Db        711 DGQLYQVVNSEVNTLKDRLDEIYKSAGKQEGCRMTFIIVSKRINSRYF--TGHR--NPVP 766

Qy        763 GTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYN 822
              |||:|  :| || ||||:|||||| |:|||| |||| |: ||  | :| |:||  |:|||
Db        767 GTVVDDVITLPERYDFFLVSQAVRIGTVSPTSYNVISDNMGLNADKLQMLSYKMTHMYYN 826

Qy        823 WPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 862
              : | ||||| | |||||||||| :||:| |:  | |:||:|
Db        827 YSGTIRVPAVCHYAHKLAFLVAESINRAPSAGLQNQLYFL 866


RESULT 12
Q9VKM1
ID   Q9VKM1      PRELIMINARY;      PRT;   843 AA.
AC   Q9VKM1; O96674; O96675;
```

```
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Piwi protein.
GN   PIWI OR CG6122.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   STRAIN=OREGON-R, AND CANTON-S;
RX   MEDLINE=99069219; PubMed=9851978;
RA   Cox D.N., Chao A., Baker J., Chang L., Qiao D., Lin H.;
RT   "A novel class of evolutionarily conserved genes defined by piwi are
RT   essential for stem cell self-renewal.";
RL   Genes Dev. 12:3715-3727(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: MEDIATES A SOMATIC SIGNALING MECHANISM REQUIRED FOR THE
CC       ASYMMETRIC DIVISION OF GERM-LINE STEM CELLS TO PRODUCE AND
CC       MAINTAIN A DAUGHTER GERM-LINE STEM CELL. IT IS NOT ESSENTIAL FOR
CC       THE FURTHER DIFFERENTIATION OF THE COMMITTED DAUGHTER CELL.
CC   -!- TISSUE SPECIFICITY: EXPRESSED SOMATICALLY IN THE OVARIOLE TERMINAL
CC       FILAMENT AND EPITHELIAL SHEATH CELLS. EXPRESSED IN THE GERMARIUM
CC       AT LOW LEVELS DURING OOGENESIS STAGES 1-6, AT A LOWER LEVEL DURING
CC       STAGES 7-9, STRONGLY AT STAGE 10 AND EVENTUALLY ACCUMULATING IN
CC       EARLY EMBRYOS.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED BOTH MATERNALLY AND ZYGOTICALLY.
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
DR   EMBL; AF104354; AAD08704.1; -.
DR   EMBL; AF104355; AAD08705.1; -.
DR   EMBL; AE003630; AAF53043.1; -.
DR   FlyBase; FBgn0004872; piwi.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
KW   Developmental protein.
FT   CONFLICT     67     67       T -> P (IN REF. 1; AAD08704).
FT   CONFLICT     84     84       A -> V (IN REF. 1; AAD08704).
FT   CONFLICT    122    124       EPS -> VPT (IN REF. 1; AAD08704).
FT   CONFLICT    576    577       NR -> KPY (IN REF. 1; AAD08705).
SQ   SEQUENCE   843 AA;  97177 MW;  9B4A95E688E9D9B7 CRC64;

  Query Match             30.9%;  Score 1398.5;  DB 5;  Length 843;
  Best Local Similarity   35.9%;  Pred. No. 2.1e-96;
  Matches  307;  Conservative  172;  Mismatches  318;  Indels   59;  Gaps   18;

Qy         48 GRGRQRGMVVGATSKSQ--------------------ELQISAGFQELSLAE----RGG 82
              ||||:| :   :| |:                     : :| |  :  :| |     ||
Db          6 GRGRRRPLNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG 65

Qy         83 --RRRDFHD--XGVNTRQNLDHVKESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYN 138
                 |: : |    :|||         || |: |: | | || ||| ::|:| :  ||:::
Db         66 PTERKPWGDQYDYLNTR---PAELVSKKGTDGVPVMLQTNFFRLKTKPEWRIVHYHVEFE 122

Qy        139 PLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVTEVFSQTRNGEHVRITITL 198
              | :|  |:|  :|  | :|:|  : |||  ||  :: : ::| :  :::     :|:|
Db        123 PSIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIKF 182

Qy        199 TNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSIL 258
                :    |  ||  |:  || :| :||: :||| ::|   |:|   :: :|||: |||
Db        183 VGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSIR 242

Qy        259 QYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTY 318
              |:| :|:| |:::|||:|:||: | |       |  |::|   :: ||||| |||:||
Db        243 QHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRTY 300

Qy        319 RVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLP 378
              |::|:|: | |||||    | ::|||:|||  :||  | |  ||:|:|:  | :
Db        301 RINDVDFGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNAS 358

Qy        379 GPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQ-- 436
                : ||||| :|||  :||:|:| :|: :: :||: |:||     ||  : |:  |
Db        359 ELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPES 415

Qy        437 -RELRDWGLSFDSNLLSFSGRILQSEKI--HQGGKTFDYNPQFADWSKETRGAPLISVKP 493
               : |||| :  | |:    |||:  : |  | |       |   ||| :  |   :::
Db        416 VKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTPS 472

Qy        494 --LDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKK-AIMIEVDDRTEAYLRALQQKV 550
                || | :|  :||     :|: :|::    ||::::      |  |||| |:||:   |
Db        473 DGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDCV 532

Qy        551 TSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMN 610
               || ::::||: ::  ::| :|||     |   |:| |  :|  | :::|:|||||:|:|
Db        533 RSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQLN 591

Qy        611 CKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASIN-EGMTRWFSRCVFQDRG 669
              ||:|   | ::: |   | :| |     |   :|:    :||:: :  : :||
Db        592 CKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSAF 651

Qy        670 QELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVG 729
              | : |   :  ||| :   :  :|||:: ||||  | || |  :||   :: ||:
Db        652 DVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKTEY 711

Qy        730 RGYN---PRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVR 786
                    |:|  ||| : :| |||       ||| |||::|  :| || |||::||| ||
Db        712 ARVQLSPPQLAYIVVTRSMNTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVSQQVR 767

Qy        787 SGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVGQS 846
              |:|||| |||:| | || |: :|:|||| ||:|||| |  |||| |||| ||| ||| :
Db        768 QGTVSPTSYNVLYSSMGLSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATLVGTN 827
```

```
Qy        847 IHREPNLSLSNRLYYL 862
              :|  |  :|  : |||
Db        828 LHSIPQNALEKKFYYL 843


RESULT 13
Q99MV6
ID   Q99MV6      PRELIMINARY;      PRT;   580 AA.
AC   Q99MV6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   PIWIL2.
GN   PIWIL1L OR PIWIL2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE=21175748; PubMed=11279525;
RA   Wang P.J., McCarrey J.R., Yang F., Page D.C.;
RT   "An abundance of X-linked genes expressed in spermatogonia.";
RL   Nat. Genet. 27:422-426(2001).
DR   EMBL; AF285586; AAK31965.1; -.
DR   MGD; MGI:1930036; Piwil2.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   580 AA;  66549 MW;  A4EA9DE8626FA136 CRC64;

  Query Match             30.0%;  Score 1360.5;  DB 11;  Length 580;
  Best Local Similarity   43.3%;  Pred. No. 8.7e-94;
  Matches  253;  Conservative  121;  Mismatches  201;  Indels    9;  Gaps    8;

Qy        284 MFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSF 343
              |  :||| :|| ||:: || |:| ||:|:|||:|||:||:||:: || :|  :|| |::|
Db          1 MHAIYQQNKEH-FQDECSKLLVGSIVITRYNNRTYRIDDVDWNKTPKDSFVMSDGKEITF 59

Qy        344 LEYYRKQYNQEITDLKQPVLVSQP-KRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFN 402
              |||| | |   : :  ||:|: :| :|:   |  | |  : :||| ::||: :|| ||
Db         60 LEYYSKNYGITVKEDDQPLLIHRPSERQNNHGMLLKGEILLLPELSFMTGIPEKMKKDFR 119

Qy        403 VMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEK 462
              ||||     |:|:|    :  |:  | :::    ||  ||||   ::    ||:|  |:
Db        120 AMKDLTQQINLSPKQHHGALECLLQRISQNETASNELTRWGLSLHKDVHKIEGRLLPMER 179

Qy        463 IHQGGKTFDYNPQFADWSKE-TRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTP 521
              |:    :|    :  :| || || | :::: |:  | | | :|  : |  |:  | |:
Db        180 INLRNTSF-VTSEGLNWVKEVTRDASILTI-PMHFWALFYPKRAMDQARELVNMLEKIAG 237

Qy        522 AMGIQMKKAIMIEV-DDRTEAYLRALQQ--KVTSDTQIVVCLLSSNRKDKYDAIKKYLCT 578
              :|::      :|: ||| | |:| :|    |    |:|||::   | | | |||| |
Db        238 PIGMRTSPPAWVELKDDRIETYIRTIQSLLGVEGKIQMVVCIIMGTRDDLYGAIKKLCCV 297

Qy        579 DCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDT 638
              | ||| :  ||:|:   : ::| || |||||:|||| ||: ||  |::|:| |||
Db        298 QSPVPSQVINVRTIGQPTRLRSVAQKILLQMNCKLGGELWGVDIPLKQLMVIGMDVYHDP 357

Qy        639 TAGRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRV 698
              : | ||: |||||||  :|:|:||| |||   ||:|| ||:|| :|: :   |  :| ::
Db        358 SRGMRSVVGFVASINLTLTKWYSRVVFQMPHQEIVDSLKLCLVGSLKKYYEVNHCLPEKI 417

Qy        699 IVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQ 758
              :|||||| |||||: |||:||   | ::    |:|:: | ||:|:::   :  :
Db        418 VVYRDGVSDGQLKTVANYEIPQLQKCFEAFD-NYHPKMVVFVVQKKISTNLYLAAPDHFV 476

Qy        759 NPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCH 818
              | ||||:|  :|  || ||:::: || |   ||||  : ::: | |||:||||:| ||
Db        477 TPSPGTVVDHTITSCEWVDFYLLAHHVRQGCGIPTHYICVLNTANLSPDHMQRLTFKLCH 536
```

```
Qy        819 VYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 862
              :|:|||| |||||||:|||||||| || :| || : |   |::|
Db        537 MYWNWPGTIRVPAPCKYAHKLAFLSGQILHHEPAIQLCGNLFFL 580


RESULT 14
P90786
ID   P90786      PRELIMINARY;      PRT;   824 AA.
AC   P90786;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   D2030.6 protein.
GN   D2030.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wilkinson J.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z73906; CAA98113.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   824 AA;  93844 MW;  970263CF6FEDBCEE CRC64;

  Query Match             28.5%;  Score 1290;  DB 5;  Length 824;
  Best Local Similarity   33.7%;  Pred. No. 3e-88;
  Matches  289;  Conservative  176;  Mismatches  311;  Indels   82;  Gaps   19;

Qy         48 GRGRQRGMVVGATSKSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTG 107
              |||| ||     :   :  :     :| : ::|                  :||||
Db          6 GRGRGRGSGSNNSGGKDQKYLGTIQPDLFIRQQG--------------------QSKTG 44

Qy        108 SSGIIVKLSTNHFRL-TSRPQWALYQYHIDYNPLMEARRLRSALLFQHE--DLIGRCHAF 164
              |||  |  |   :  ::  :::|||:::  | ::::  |  :|  :   | ||    |
Db         45 SSGQPQKCFANFIPIEMTQSDYSIYQYHVEFEPTVDSKANREKMLRDNNVTDEIGHHFVF 104

Qy        165 DGTILFLPKRL-QHKVTEVFSQTRNGEH------VRITITLTNELPPTSPTCLQFYNIXF 217
              || ||:|  :   |::: ||       :|      : |    ||       |  :  :|
Db        105 DGMILYLKEEWEQNQMIEV-------QHPIDRSLICIRFKQTNRFLVDDPQTINIFNTII 157

Qy        218 RRLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRS 277
              ||    : | |:|||||:|  |   :|:: : | ||: |:|  || | ||| :   |::|
Db        158 RRSFDALQLTQLGRNYFNWGDSRAVPDYNMSILPGYETAIRMYEENFMLCVENRFKMVRE 217

Qy        278 ETVLDFMFNLYQ------QTEEHKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKS 331
              |:    |: |:       |    : ||::::   |  ::|:|||| :|   :|:  :| |
Db        218 ES----MYILFHKELRSCQNNPQRVQEKMNEMYGGTTIITRYNNKLHRYTRLDYSISPLS 273

Qy        332 TFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMXIPELCYLT 391
              | | ||  ::  ||:: ||  |||   ||:::|: | ::   |  |  :  :||||: |
Db        274 EFVK-DGQSITLKEYFKNQYGIEITVDDQPIIISEGKPKQ--PGEPPQVSYIVPELCFPT 330

Qy        392 GLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLL 451
              ||||:|| || :||::| |||::|:||  |  :||  : |:: |    : ||:|   :|
Db        331 GLTDEMRKDFKMMKEIAKHTRMSPQQRLVESRKLIVDLSKNEKVMECFKYWGISLGQDLA 390

Qy        452 SFSGRILQSEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEA--A 509
              :  |:|:|| : || ||  |  : |:|::  :  :     : ||::|           :
Db        391 NVQARVLKSEPL-QGKKT--YEGKQAEWARGVKECGIYRGSNMTNWIVIGPGSGNSGLLS 447

Qy        510 NSLIQNLFKVTPAMGIQMKKAIMIEV-----DDRTEAYLRALQQKVTSDTQIVVCLLSSN 564
```

```
Db        448 QKFIEEARRLGKILQVQLGEPMCVPIRGISPNDYLEGVKGAIKQVDGEDIHMLVVMLADD 507

Qy        565 RKDKYDAIKKYLCTDCPTPSQCVVARTL-------GKQQTVMAIATKIALQMNCKMGGEL 617

Db        508 NKTRYDSLKKFLCVECPIPNQCVNLRTLAGKSKDGGENKNLGSIVLKIVLQMICKTGGAL 567

Qy        618 WRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLK 677

Db        568 WKVNIPLKNTMIVGYDLYHDSTLKGKTVGACVSTTSNDFTQFYSQTRPHENPTQLGNNLT 627

Qy        678 VCLQAALRAWSGCNEY-MPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGR------ 730

Db        628 HFVRKALKQYYDSNDQTLPSRLILYRDGAGDGQIPYIKNTEVKLVRDACDAVTDKAAELS 687

Qy        731 ---GYNPRLTVIVVKKRVNARFFAQSGGRLQ---NPLPGTVIDVEVTRPEWYDFFIVSQA 784

Db        688 NKVQEKIKLAFIIVTKRVNMRILKQ-GSSLDNAINPQPGTVVDTTVTRPERMDFYLVPQF 746

Qy        785 VRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVG 844

Db        747 VNQGTVTPVSYNIIHDDTDLGPDKHQQLAFKLCHLYYNWQGTVRVPAPCQYAHKLAFLTA 806

Qy        845 QSIHREPNLSLSNRLYYL 862

Db        807 QSLHDDANGCLRDKLFFL 824
```

RESULT 15
Q96SW6
ID Q96SW6 PRELIMINARY; PRT; 530 AA.
AC Q96SW6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE CDNA FLJ14591 fis, clone NT2RM4002034, weakly similar to Homo
DE sapiens hiwi mRNA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AK027497; BAB55155.1; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02171; Piwi; 1.
SQ SEQUENCE 530 AA; 60620 MW; BA4A8AADD809B32F CRC64;

Query Match 26.3%; Score 1190; DB 4; Length 530;
Best Local Similarity 42.2%; Pred. No. 5.1e-81;
Matches 224; Conservative 103; Mismatches 198; Indels 6; Gaps 5;

```
Qy        336 ADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRR-GPGGTLPGPAMXIPELCYLTGLT 394

Db          2 SDGKEITFLEYYSKNYGITVKEEDQPLLIHRPSERQDNHGMLLKGEILLLPELSFMTGIP 61

Qy        395 DKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFS 454

Db         62 EKMKKDFRAMKDLAQQINLSPKQHHSALECLLQRIAKNEAATNELMRWGLRLQKDVHKIE 121

Qy        455 GRILQSEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQ 514

Db        122 GRVLPMERINLKNTSF-ITSQELNWVKEVTRDPSILTIPMHFWALFYPKRAMDQARELVN 180

Qy        515 NLFKVTPAMGIQMKKAIMIEV-DDRTEAYLRALQQKVTSD--TQIVVCLLSSNRKDKYDA 571

Db        181 MLEKIAGPIGMRMSPPAWVELKDDRIETYVRTIQSTLGAEGKIQMVVCIIMGPRDDLYGA 240

Qy        572 IKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMALKLAMIVG 631

Db        241 IKKLCCVQSPVPSQVVNVRTIGQPTRLRSVAQKILLQINCKLGGELWGVDIPLKQLMVIG 300

Qy        632 IDCYHDTTAGRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCN 691

Db        301 MDVYHDPSRGMRSVVGFVASINLTLTKWYSRVVFQMPHQEIVDSLKLCLVGSLKKFYEVN 360

Qy        692 EYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIVVKKRVNARFFA 751

Db        361 HCLPEKIVVYRDGVSDGQLKTVANYEIPQLQKCFEAF-ENYQPKMVVFVVQKKISTNLYL 419

Qy        752 QSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQR 811

Db        420 AAPQNFVTPTPGTVVDHTITSCEWVDFYLLAHHVRQGCGIPTHYVCVLNTANLSPDHMQR 479

Qy        812 LTYKXCHVYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 862

Db        480 LTFKLCHMYWNWPGTIRVPAPCKYAHKLAFLSGHILHHGPAIQLCENLFFL 530
```

Search completed: July 22, 2003, 07:04:48
Job time : 34.5931 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2003, 18:02:32 ; Search time 8751.36 Seconds
(without alignments)
11546.204 Million cell updates/sec

Title: US-09-873-737A-5
Perfect score: 3472
Sequence: 1 gggcgaattgggcccgacgt.....taaaataaaattgatgtccc 3472

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		3359	96.7	3362	9	AF104260	AF104260 Homo sapi
2		3344.6	96.3	3399	9	BC028581	BC028581 Homo sapi
3		3337.6	96.1	3421	9	AF387507	AF387507 Homo sapi
4		2513.6	72.4	2579	9	AK093133	AK093133 Homo sapi
5		2311	66.6	2328	9	AF264004	AF264004 Homo sapi
6		1839.2	53.0	3910	10	AB032604	AB032604 Mus muscu
7		1839.2	53.0	4064	10	AF438405	AF438405 Mus muscu
8		1067.2	30.7	3273	5	AF336369	AF336369 Danio rer
9	c	830.8	23.9	193697	2	AC025837	AC025837 Homo sapi
10		830.8	23.9	198104	2	AC127071	AC127071 Homo sapi
11		679.2	19.6	3006	9	BC031060	BC031060 Homo sapi
12		679.2	19.6	3138	9	AK093475	AK093475 Homo sapi
13		613.2	17.7	2456	9	AK096837	AK096837 Homo sapi
14		565	16.3	3057	3	AY014899	AY014899 Strongylo
15	c	550.4	15.9	590	6	AX381696	AX381696 Sequence
16		495.8	14.3	2605	3	AY014900	AY014900 Strongylo
17		459.4	13.2	501	6	AX395819	AX395819 Sequence
18	c	458.4	13.2	460	6	AX209866	AX209866 Sequence
19		456.6	13.2	4913	10	AB032605	AB032605 Mus muscu
20		438.8	12.6	3610	9	BC025995	BC025995 Homo sapi
21		434.2	12.5	3620	9	AK056418	AK056418 Homo sapi
22		378.4	10.9	1723	3	AY014901	AY014901 Strongylo
23		377	10.9	3649	6	AX247684	AX247684 Sequence
24		377	10.9	3659	10	AF285586	AF285586 Mus muscu
25		329	9.5	2272	9	AK027497	AK027497 Homo sapi
26		310.2	8.9	2761	3	AF145680	AF145680 Drosophil
27		253.4	7.3	3047	3	AF104354	AF104354 Drosophil
28		238.8	6.9	1733	9	AK001213	AK001213 Homo sapi
29		203.4	5.9	61995	2	AC090565	AC090565 Homo sapi
30		186.2	5.4	7781	3	DMDNARPL9	X94613 D.melanogas
31	c	184.6	5.3	37980	3	AC005116	AC005116 Drosophil
32	c	184.6	5.3	78507	2	AC019704	AC019704 Drosophil
33	c	184.6	5.3	169210	3	AC092226	AC092226 Drosophil
34	c	184.6	5.3	176306	3	AC007186	AC007186 Drosophil
35	c	184.6	5.3	276576	3	AE003630	AE003630 Drosophil
36		183	5.3	901	3	AF334408	AF334408 Drosophil
37		170.8	4.9	548	6	AX387649	AX387649 Sequence
38		170.6	4.9	236516	2	AC111089	AC111089 Mus muscu
39	c	160.4	4.6	3677	2	AC017885	AC017885 Drosophil
40		160.4	4.6	3855	3	AE003107	AE003107 Drosophil
41		141.2	4.1	9359	3	AF104355	AF104355 Drosophil
42	c	138.4	4.0	39365	2	AC100141	AC100141 Mus muscu
43	c	129.6	3.7	66611	2	AC090147	AC090147 Homo sapi
44		126.6	3.6	66611	2	AC090147	AC090147 Homo sapi
45		126.4	3.6	1263	9	AK000397	AK000397 Homo sapi

ALIGNMENTS

RESULT 1
AF104260
LOCUS AF104260 3362 bp mRNA linear PRI 09-JAN-2002
DEFINITION Homo sapiens HIWI mRNA, complete cds.
ACCESSION AF104260
VERSION AF104260.2 GI:18098557
KEYWORDS .
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3362)
AUTHORS Cox,D.N., Chao,A., Baker,J., Chang,L., Qiao,D. and Lin,H.
TITLE A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal

```
  JOURNAL   Genes Dev. 12 (23), 3715-3727 (1998)
  MEDLINE   99069219
   PUBMED   9851978
REFERENCE   2  (bases 1 to 3362)
  AUTHORS   Qiao,D., Zeeman,A.-M., Deng,W., Looijenga,L.H.J. and Lin,H.
  TITLE     Molecular characterization of hiwi, a human member of the piwi stem
            cell gene family whose overexpression is correlated to seminomas
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 3362)
  AUTHORS   Lin,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1998) Cell Biology, Duke University Medical
            Center, 412 Nanaline Duke Bldg., Research Dr., Durham, NC 27710,
            USA
REFERENCE   4  (bases 1 to 3362)
  AUTHORS   Qiao,D., Zeeman,A.-M., Deng,W., Looijenga,L.H.J. and Lin,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Cell Biology, Duke University, 412 Nanaline
            Duke Building, DUMC, Durham, NC 27710, USA
  REMARK    Nucleotide sequence updated by submitter
COMMENT     On Jan 9, 2002 this sequence version replaced gi:4038412.
FEATURES             Location/Qualifiers
     source          1. .3362
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /map="12q24.33"
                     /tissue_type="testis"
     CDS             57. .2642
                     /note="member of the PIWI family"
                     /codon_start=1
                     /product="HIWI"
                     /protein_id="AAC97371.2"
                     /db_xref="GI:18098558"
                     /translation="MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEG
                     ELFGRGRQRGTAGGTAKSQGLQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVKE
                     SKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKC
                     HAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRR
                     LLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRS
                     ETVLDFMFNFYHQTEEHKFQEQVSKELIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKK
                     ADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGL
                     TDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLL
                     SFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA
                     NSLIQNLFKVTPAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKD
                     KYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLK
                     LVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAA
                     LRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVV
                     KKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIY
                     DNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSN
                     RLYYL"
BASE COUNT     1026 a    659 c    759 g    918 t
ORIGIN

  Query Match             96.7%;  Score 3359;  DB 9;  Length 3362;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 3359;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        111 CTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAACAATGA 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAACAATGA 60

Qy        171 CTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGCTGG 230
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGCTGG 120

Qy        231 TGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGCCACCAG 290
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGCCACCAG 180

Qy        291 CAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAGCCAAGT 350
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAGCCAAGT 240

Qy        351 CACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAGAGAGGAGGTC 410
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        241 CACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTAGCAGAGAGAGGAGGTC 300

Qy        411 GTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGTTAAAG 470
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGTTAAAG 360

Qy        471 AATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGA 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGA 420

Qy        531 CATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGGAAG 590
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGGAAG 480

Qy        591 CCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCATG 650
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCATG 540

Qy        651 CTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTTT 710
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTTT 600

Qy        711 TTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATGAACTTC 770
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATGAACTTC 660

Qy        771 CACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAAA 830
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAAA 720

Qy        831 TCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCAATTGATATTC 890
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCAATTGATATTC 780

Qy        891 CAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAACA 950
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAACA 840

Qy        951 GCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGATT 1010
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGATT 900

Qy       1011 TCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAAG 1070
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAAG 960

Qy       1071 AANTAATAGGTTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGAGTGGATGATA 1130
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 AACTAATAGGTTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGAGTGGATGATA 1020

Qy       1131 TTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGGGGTCAGCT 1190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGGGGTCAGCT 1080

Qy       1191 TCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTCT 1250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTCT 1140

Qy       1251 TGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATGC 1310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATGC 1200

Qy       1311 TCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAACG 1370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAACG 1260

Qy       1371 TGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTGG 1430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTGG 1320

Qy       1431 GACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGG 1490
```

```
Db      1321 GACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGG 1380
Qy      1491 GTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAGA 1550
Db      1381 GTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAGA 1440
Qy      1551 TTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAA 1610
Db      1441 TTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAA 1500
Qy      1611 CAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACGC 1670
Db      1501 CAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACGC 1560
Qy      1671 GAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCA 1730
Db      1561 GAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCA 1620
Qy      1731 TGGGCATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACTGAAGCCTACT 1790
Db      1621 TGGGCATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACTGAAGCCTACT 1680
Qy      1791 TAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCTGTTGTCAA 1850
Db      1681 TAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCTGTTGTCAA 1740
Qy      1851 GTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCC 1910
Db      1741 GTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCC 1800
Qy      1911 CAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTA 1970
Db      1801 CAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTA 1860
Qy      1971 CAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCC 2030
Db      1861 CAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCC 1920
Qy      2031 CCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGA 2090
Db      1921 CCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGA 1980
Qy      2091 GGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACGCT 2150
Db      1981 GGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACGCT 2040
Qy      2151 GCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGG 2210
Db      2041 GCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGG 2100
Qy      2211 CTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATCGTGTACCGCG 2270
Db      2101 CTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATCGTGTACCGCG 2160
Qy      2271 ATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCACAGTTTTTGG 2330
Db      2161 ATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCACAGTTTTTGG 2220
Qy      2331 ATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATTGTGGTGAAGA 2390
Db      2221 ATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGACTAACGGTAATTGTGGTGAAGA 2280
Qy      2391 AAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTCCTG 2450
Db      2281 AAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTCCTG 2340
Qy      2451 GAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTTATCGTGAGCC 2510
Db      2341 GAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTTATCGTGAGCC 2400
Qy      2511 AGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTATGACAACAGCG 2570
Db      2401 AGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTATGACAACAGCG 2460
Qy      2571 GCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATCTATTACAACT 2630
Db      2461 GCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATCTATTACAACT 2520
Qy      2631 GGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTG 2690
Db      2521 GGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTG 2580
Qy      2691 TTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCT 2750
Db      2581 TTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCT 2640
Qy      2751 AACCTGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAG 2810
Db      2641 AACCTGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAG 2700
Qy      2811 CTTTTATTTACTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAG 2870
Db      2701 CTTTTATTTACTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAG 2760
Qy      2871 GAGATCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATATGTAAA 2930
Db      2761 GAGATCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATATGTAAA 2820
Qy      2931 AATTAAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTG 2990
Db      2821 AATTAAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTG 2880
Qy      2991 TTTATTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTA 3050
Db      2881 TTTATTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTA 2940
Qy      3051 TTTGAAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTA 3110
Db      2941 TTTGAAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTA 3000
Qy      3111 TAAACCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTTAAAGGAGATAAGAGGC 3170
Db      3001 TAAACCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTTAAAGGAGATAAGAGGC 3060
Qy      3171 GTAAAGTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTCATATCTTAAAGATAAA 3230
Db      3061 GTAAAGTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTCATATCTTAAAGATAAA 3120
Qy      3231 AGGTACTATTATATAACCTATACACAAGATACAGGAGAAAATATGCTTGATTTTTATTTG 3290
Db      3121 AGGTACTATTATATAACCTATACACAAGATACAGGAGAAAATATGCTTGATTTTTATTTG 3180
Qy      3291 GCAGGGGGGCTAGGTTGTATGGGAGTAAAAAAAACATTGAAAATTTTTAAATTGTCCAAA 3350
Db      3181 GCAGGGGGGCTAGGTTGTATGGGAGTAAAAAAAACATTGAAAATTTTTAAATTGTCCAAA 3240
Qy      3351 GAAACATTTTAAGACTCTTTAACAAAAAAGGCCATGAGTAAATCTCTATATTAACATCAC 3410
Db      3241 GAAACATTTTAAGACTCTTTAACAAAAAAGGCCATGAGTAAATCTCTATATTAACATCAC 3300
Qy      3411 TATTTATTTTGTTTTGGAACTGGGACATGATTCTATTTGTTATAAAATAAAATTGATGTC 3470
Db      3301 TATTTATTTTGTTTTGGAACTGGGACATGATTCTATTTGTTATAAAATAAAATTGATGTC 3360
Qy      3471 CC 3472
Db      3361 CC 3362
```

RESULT 2
BC028581
LOCUS BC028581 3399 bp mRNA linear PRI 25-APR-2002
DEFINITION Homo sapiens, piwi-like 1 (Drosophila), clone MGC:26748 IMAGE:4827445, mRNA, complete cds.
ACCESSION BC028581

```
VERSION     BC028581.1  GI:20306781
KEYWORDS    MGC.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3399)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 34 Row: b Column: 20
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, Similarity but not identity to protein.
FEATURES             Location/Qualifiers
     source          1. .3399
                     /organism="Homo sapiens"
                     /db_xref="LocusID:9271"
                     /db_xref="taxon:9606"
                     /clone="MGC:26748 IMAGE:4827445"
                     /tissue_type="Testis"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
     CDS             74. .2659
                     /codon_start=1
                     /product="piwi-like 1 (Drosophila)"
                     /protein_id="AAH28581.1"
                     /db_xref="GI:20306782"
                     /translation="MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEG
                     ELFGRGRQRGTAGGTAKSQGLQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVKE
                     SKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKC
                     HAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRR
                     LLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRS
                     ETVLDFMFNFYHQTEEHKFQEQVSKELIGLVVLTKYNNKTYRVDDIDWDQNPKSTFKK
                     ADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGL
                     TDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLL
                     SFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVNPLDNWLLIYTRRNYEAA
                     NSLIQNLFKVTPAMGMQMKKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKD
                     KYDAIKKYPCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLK
                     LVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAA
                     LRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVV
                     KKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIY
                     DNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSN
                     RLYYL"
BASE COUNT     1048 a    662 c    766 g    923 t
ORIGIN

  Query Match          96.3%;  Score 3344.6;  DB 9;  Length 3399;
  Best Local Similarity  99.6%;  Pred. No. 0;
  Matches 3350;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        108 GGTCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAACAA 167
              || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         15 GGGCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAACAA 74

Qy        168 TGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGC 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         75 TGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGC 134

Qy        228 TGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGCCAC 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 TGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGCCAC 194

Qy        288 CAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAGCCA 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 CAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAGCCA 254

Qy        348 AGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAGAGAGGAG 407
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        255 AGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTAGCAGAGAGAGGAG 314

Qy        408 GTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGTTA 467
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        315 GTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGTTA 374

Qy        468 AAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGC 527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        375 AAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGC 434

Qy        528 TGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGG 587
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        435 TGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGG 494

Qy        588 AAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTC 647
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        495 AAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTC 554

Qy        648 ATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAG 707
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        555 ATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAG 614

Qy        708 TTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATGAAC 767
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        615 TTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATGAAC 674

Qy        768 TTCCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGA 827
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        675 TTCCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGA 734

Qy        828 AAATCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCAATTGATA 887
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        735 AAATCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCAATTGATA 794

Qy        888 TTCCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAA 947
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        795 TTCCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAA 854

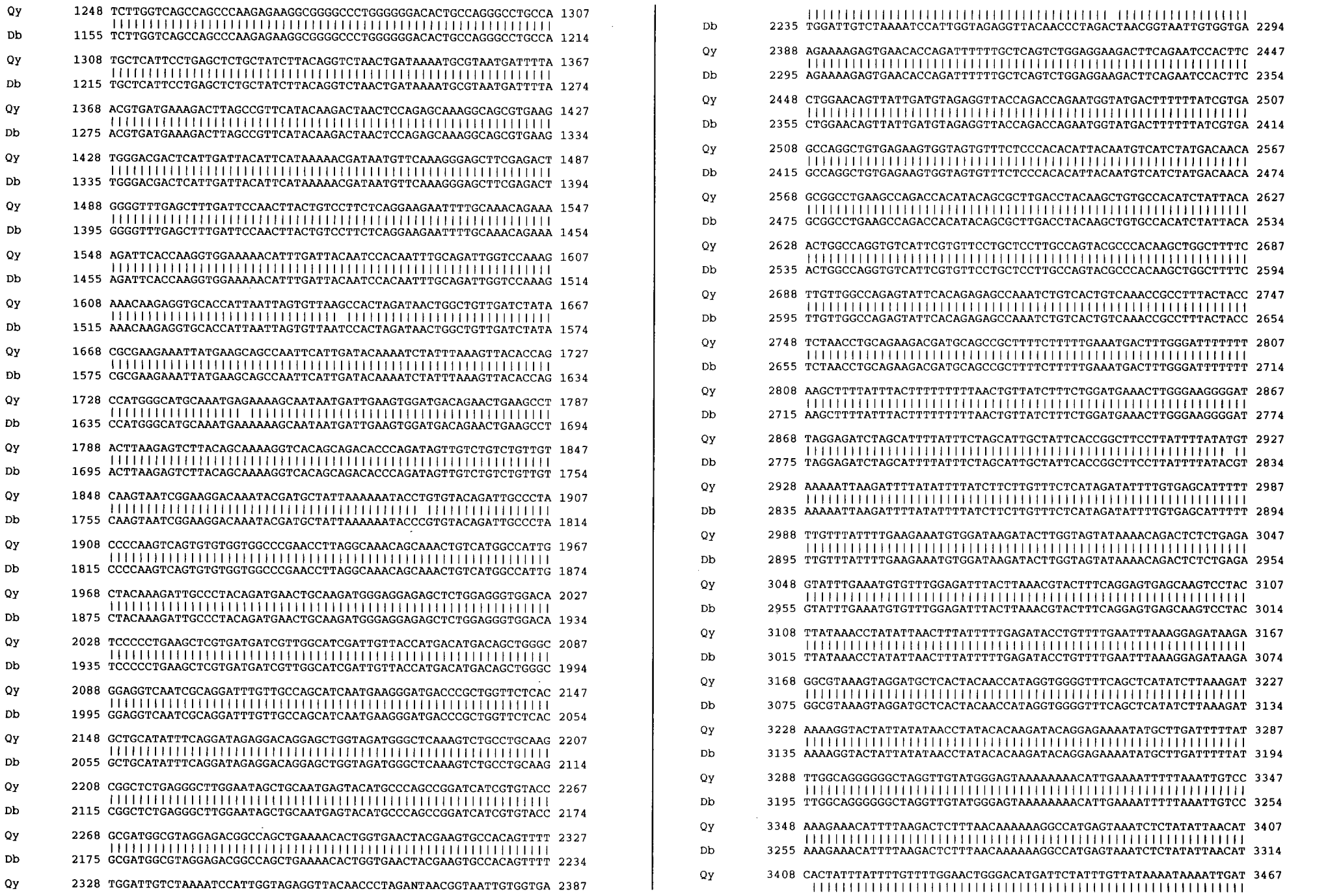
Qy        948 ACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGG 1007
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 ACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGG 914

Qy       1008 ATTTCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAAGTTTCCA 1067
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        915 ATTTCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAAGTTTCCA 974

Qy       1068 AAGAANTAATAGGTTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGAGTGGATG 1127
              ||||| |||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        975 AAGAACTAATAGGTTTAGTTGTTCTTACCAAGTATAACAATAAGACATACAGAGTGGATG 1034

Qy       1128 ATATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGGGGTCA 1187
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       1035 ATATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGAAGTCA 1094

Qy       1188 GCTTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTG 1247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1095 GCTTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTG 1154
```


Qy 1248 TCTTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCA 1307
Db 1155 TCTTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCA 1214
Qy 1308 TGCTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTA 1367
Db 1215 TGCTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTA 1274
Qy 1368 ACGTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAG 1427
Db 1275 ACGTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAG 1334
Qy 1428 TGGGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACT 1487
Db 1335 TGGGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACT 1394
Qy 1488 GGGGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAA 1547
Db 1395 GGGGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAA 1454
Qy 1548 AGATTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCCAAAG 1607
Db 1455 AGATTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCCAAAG 1514
Qy 1608 AAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATA 1667
Db 1515 AAACAAGAGGTGCACCATTAATTAGTGTTAATCCACTAGATAACTGGCTGTTGATCTATA 1574
Qy 1668 CGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACCAG 1727
Db 1575 CGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACCAG 1634
Qy 1728 CCATGGGCATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACTGAAGCCT 1787
Db 1635 CCATGGGCATGCAAATGAAAAAAGCAATAATGATTGAAGTGGATGACAGAACTGAAGCCT 1694
Qy 1788 ACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCTGTTGT 1847
Db 1695 ACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCTGTTGT 1754
Qy 1848 CAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTGCCCTA 1907
Db 1755 CAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCCGTGTACAGATTGCCCTA 1814
Qy 1908 CCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTG 1967
Db 1815 CCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTG 1874
Qy 1968 CTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACA 2027
Db 1875 CTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACA 1934
Qy 2028 TCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGC 2087
Db 1935 TCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGC 1994
Qy 2088 GGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCAC 2147
Db 1995 GGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCAC 2054
Qy 2148 GCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAG 2207
Db 2055 GCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAG 2114
Qy 2208 CGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATCGTGTACC 2267
Db 2115 CGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATCGTGTACC 2174
Qy 2268 GCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCACAGTTTT 2327
Db 2175 GCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCACAGTTTT 2234
Qy 2328 TGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATTGTGGTGA 2387
Db 2235 TGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGACTAACGGTAATTGTGGTGA 2294
Qy 2388 AGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTC 2447
Db 2295 AGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTC 2354
Qy 2448 CTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTTATCGTGA 2507
Db 2355 CTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTTATCGTGA 2414
Qy 2508 GCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTATGACAACA 2567
Db 2415 GCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTATGACAACA 2474
Qy 2568 GCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATCTATTACA 2627
Db 2475 GCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATCTATTACA 2534
Qy 2628 ACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTC 2687
Db 2535 ACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTC 2594
Qy 2688 TTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACC 2747
Db 2595 TTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACC 2654
Qy 2748 TCTAACCTGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTT 2807
Db 2655 TCTAACCTGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTT 2714
Qy 2808 AAGCTTTTATTTACTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGAT 2867
Db 2715 AAGCTTTTATTTACTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGAT 2774
Qy 2868 TAGGAGATCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATATGT 2927
Db 2775 TAGGAGATCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATACGT 2834
Qy 2928 AAAAATTAAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTT 2987
Db 2835 AAAAATTAAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTT 2894
Qy 2988 TTGTTTATTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGA 3047
Db 2895 TTGTTTATTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGA 2954
Qy 3048 GTATTTGAAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTAC 3107
Db 2955 GTATTTGAAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTAC 3014
Qy 3108 TTATAAACCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTTAAAGGAGATAAGA 3167
Db 3015 TTATAAACCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTTAAAGGAGATAAGA 3074
Qy 3168 GGCGTAAAGTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTCATATCTTAAAGAT 3227
Db 3075 GGCGTAAAGTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTCATATCTTAAAGAT 3134
Qy 3228 AAAAGGTACTATTATATAACCTATACACAAGATACAGGAGAAAATATGCTTGATTTTTAT 3287
Db 3135 AAAAGGTACTATTATATAACCTATACACAAGATACAGGAGAAAATATGCTTGATTTTTAT 3194
Qy 3288 TTGGCAGGGGGGCTAGGTTGTATGGGAGTAAAAAAAACATTGAAAATTTTTAAATTGTCC 3347
Db 3195 TTGGCAGGGGGGCTAGGTTGTATGGGAGTAAAAAAAACATTGAAAATTTTTAAATTGTCC 3254
Qy 3348 AAAGAAACATTTTAAGACTCTTTAACAAAAAAGGCCATGAGTAAATCTCTATATTAACAT 3407
Db 3255 AAAGAAACATTTTAAGACTCTTTAACAAAAAAGGCCATGAGTAAATCTCTATATTAACAT 3314
Qy 3408 CACTATTTATTTTGTTTTGGAACTGGGACATGATTCTATTTGTTATAAAATAAAATTGAT 3467

```
Db       3315 TACTATTTATTTTGTTTTGGAACTGGGACATGATTCTATTTGTTATAAAATAAAATTGAT 3374

Qy       3468 GT 3469
              ||
Db       3375 GT 3376


RESULT 3
AF387507
LOCUS       AF387507                3421 bp    mRNA    linear   PRI 01-JUL-2001
DEFINITION  Homo sapiens PIWI protein mRNA, complete cds.
ACCESSION   AF387507
VERSION     AF387507.1  GI:14579644
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3421)
  AUTHORS   Sha,J.H.
  TITLE     Cloning and identification of human piwi protein related to testis
            development
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3421)
  AUTHORS   Sha,J.H., Li,J.M. and Zhou,Z.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-2001) Key Lab of Reproductive Medicine, Nanjing
            Medical University, 140 Han Zhong Road, Nanjing Medical University,
            Jiangsu 210029, China
FEATURES             Location/Qualifiers
     source          1. .3421
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="testis"
                     /dev_stage="adult"
     CDS             64. .2649
                     /note="higher expression in adult testis than 6 month old
                     embryo testis"
                     /codon_start=1
                     /product="PIWI protein"
                     /protein_id="AAK69348.1"
                     /db_xref="GI:14579645"
                     /translation="MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEG
                     ELFGRGRQRGTAGGTAKSQGLQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVKE
                     SKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKC
                     HAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRR
                     LLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRS
                     ETVLDFMFNFYHQTEEHKFQEQVSKELIGLVVLTKYNNKTYRVDDIDWDQNPKSTFKK
                     ADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGL
                     TDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLL
                     SFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA
                     NSLIQNLFKVTPAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKD
                     KYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLK
                     LVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAA
                     LRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVV
                     KKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIY
                     DNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSN
                     RLYYL"
     polyA_signal    3354. .3359
BASE COUNT     1071 a    661 c    767 g    922 t
ORIGIN

  Query Match           96.1%;  Score 3337.6;  DB 9;  Length 3421;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 3354;  Conservative    0;  Mismatches   12;  Indels    1;  Gaps    1;

Qy        104 GGCAGGTCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAA 163
              ||  || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGGGGGGCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAA 60

Qy        164 ACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCG 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCG 120

Qy        224 CAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCG 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCG 180

Qy        284 CCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACA 343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACA 240

Qy        344 GCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAGAGA 403
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        241 GCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTAGCAGAGAGA 300

Qy        404 GGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCAT 463
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCAT 360

Qy        464 GTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTC 523
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTC 420

Qy        524 CGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTG 583
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTG 480

Qy        584 ATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAG 643
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAG 540

Qy        644 TGTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACT 703
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TGCCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACT 600

Qy        704 GAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAAT 763
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAAT 660

Qy        764 GAACTTCCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTT 823
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAACTTCCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTT 720

Qy        824 TTGAAAATCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCAATT 883
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TTGAAAATCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCAATT 780

Qy        884 GATATTCCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGTAT 943
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GATATTCCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGTAT 840

Qy        944 GAAAACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTT 1003
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTT 900

Qy       1004 TTGGATTTCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAAGTT 1063
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTGGATTTCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAAGTT 960

Qy       1064 TCCAAAGAANTAATAGGTTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGAGTG 1123
              |||||||| ||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        961 TCCAAAGAACTAATAGGTTTAGTTGTTCTTACCAAGTATAACAATAAGACATACAGAGTG 1020

Qy       1124 GATGATATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGGG 1183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GATGATATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGAA 1080

Qy       1184 GTCAGCTTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAG 1243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCAGCTTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAG 1140

Qy       1244 CCTGTCTTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCT 1303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCTGTCTTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCT 1200
```

```
Qy       1304 GCCATGCTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGAT 1363
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GCCATGCTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGAT 1260

Qy       1364 TTTAACGTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGT 1423
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTTAACGTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGT 1320

Qy       1424 GAAGTGGGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGA 1483
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GAAGTGGGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGA 1380

Qy       1484 GACTGGGGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACA 1543
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GACTGGGGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACA 1440

Qy       1544 GAAAAGATTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCC 1603
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAAAAGATTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCC 1500

Qy       1604 AAAGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATC 1663
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAAGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATC 1560

Qy       1664 TATACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACA 1723
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TATACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACA 1620

Qy       1724 CCAGCCATGGGCATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACTGAA 1783
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CCAGCCATGGGCATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACTGAA 1680

Qy       1784 GCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCTG 1843
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCTG 1740

Qy       1844 TTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTGC 1903
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 TTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTGC 1800

Qy       1904 CCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCC 1963
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCC 1860

Qy       1964 ATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTG 2023
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTG 1920

Qy       2024 GACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCT 2083
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCT 1980

Qy       2084 GGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTC 2143
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTC 2040

Qy       2144 TCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTG 2203
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTG 2100

Qy       2204 CAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATCGTG 2263
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATCGTG 2160

Qy       2264 TACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCACAG 2323
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 TACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCACAG 2220

Qy       2324 TTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATTGTG 2383
             ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       2221 TTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGACTAACGGTAATTGTG 2280

Qy       2384 GTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAATCCA 2443
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 GTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAATCCA 2340

Qy       2444 CTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTTATC 2503
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTTATC 2400

Qy       2504 GTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTATGAC 2563
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTATGAC 2460

Qy       2564 AACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATCTAT 2623
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 AACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATCTAT 2520

Qy       2624 TACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCT 2683
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 TACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCT 2580

Qy       2684 TTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTAC 2743
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 TTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTAC 2640

Qy       2744 TACCTCTAACCTGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTT 2803
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 TACCTCTAACCTGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTT 2700

Qy       2804 TTTTAAGCTTTTATTTACTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGG 2863
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 TTTTAAGCTTTTATTTACTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGG 2760

Qy       2864 GGATTAGGAGATCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTAT 2923
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 GGATTAGGAGATCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTAT 2820

Qy       2924 ATGTAAAAATTAAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCAT 2983
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 ATGTAAAAATTAAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCAT 2880

Qy       2984 TTTTTTGTTTATTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCT 3043
             ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       2881 TTTTTTGTTTATTTTGAAGAAATGTGGATAAGATACTTGGTAGTATGAAACAGACTCTCT 2940

Qy       3044 GAGAGTATTTGAAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTC 3103
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 GAGAGTATTTGAAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTC 3000

Qy       3104 CTACTTATAAACCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTTAAAGGAGAT 3163
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 CTACTTATAAACCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTTAAAGGAGAT 3060

Qy       3164 AAGAGGCGTAAAGTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTCATATCTTAA 3223
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 AAGAGGCGTAAAGTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTCATATCTTAA 3120

Qy       3224 AGATAAAAGGTACTATTATATAACCTATACACAAGATACAGGAGAAAATATGCTTGATTT 3283
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 AGATAAAAGGTACTATTATATAACCTATACACAAGATACAGGAGAAAATATGCTTGATTT 3180

Qy       3284 TTATTTGGCAGGGGGGCTAGGTTGTATGGGAGT-AAAAAAAACATTGAAAATTTTTAAAT 3342
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       3181 TTATTTGGCAGGGGGGCTAGGTTGTATGGGAGTAAAAAAAAACATTGAAAATTTTTAAAT 3240

Qy       3343 TGTCCAAAGAAACATTTTAAGACTCTTTAACAAAAAAGGCCATGAGTAAATCTCTATATT 3402
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 TGTCCAAAGAAACATTTTAAGACTCTTTAACAAAAAAGGCCATGAGTAAATCTCTATATT 3300

Qy       3403 AACATCACTATTTATTTTGTTTTGGAACTGGGACATGATTCTATTTGTTATAAAATAAAA 3462
             ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 AACATTACTATTTATTTTGTTTTGGAACTGGGACATGATTCTATTTGTTATAAAATAAAA 3360

Qy       3463 TTGATGT 3469
             |||||||
```

```
Db        3361 TTGATGT 3367


RESULT 4
AK093133
LOCUS       AK093133               2579 bp    mRNA    linear   PRI 15-JUL-2002
DEFINITION  Homo sapiens cDNA FLJ35814 fis, clone TESTI2006083, highly similar
            to Mus musculus mRNA for MIWI (piwi).
ACCESSION   AK093133
VERSION     AK093133.1  GI:21751907
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens testis cDNA to mRNA, clone_lib:TESTI2
            clone:TESTI2006083.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,
            Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R.,
            Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kawakami,B.,
            Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and
            Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2579)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1. .2579
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="TESTI2006083"
                     /tissue_type="testis"
                     /clone_lib="TESTI2"
                     /note="cloning vector: pME18SFL3"
     CDS             54. .2543
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAC04068.1"
                     /db_xref="GI:21751908"
                     /translation="MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEG
                     ELFGRGRQRGTAGGTAKSQGLQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVKE
                     SKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKC
                     HAFDGTILFLPKRLQQKATEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRR
                     LLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRS
                     ETVLDFMFNFYHQTEEHKFQEQVSKELIGLVVLTKYNNKTYRVDDIDWDQNPKSTFKK
                     ADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGL
                     TDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLL
                     SFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA
                     NSLIQNLFKVTPAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKD
                     KYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLK
                     LVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAA
                     LRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVV
                     KKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIY
                     DNSGLKPDHIQRLTYKLCHIYYNWPVSASTC"
BASE COUNT      797 a    542 c    613 g    627 t
ORIGIN

  Query Match          72.4%;  Score 2513.6;  DB 9;  Length 2579;

  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 2516;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        114 AGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAACAATGACTG 173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAACAATGACTG 60

Qy        174 GGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGCTGGTGG 233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGCTGGTGG 120

Qy        234 GCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGCCACCAGCAG 293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGCCACCAGCAG 180

Qy        294 AGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAGCCAAGTCAC 353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAGCCAAGTCAC 240

Qy        354 AAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAGAGAGGAGGTCGTC 413
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db        241 AAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTAGCAGAGAGAGGAGGTCGTC 300

Qy        414 GTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGTTAAAGAAT 473
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGTTAAAGAAT 360

Qy        474 CAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGACAT 533
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGACAT 420

Qy        534 CCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGGAAGCCA 593
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGGAAGCCA 480

Qy        594 GAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCATGCTT 653
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCATGCTT 540

Qy        654 TTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTTTTTA 713
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        541 TTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGCTACTGAAGTTTTTA 600

Qy        714 GTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATGAACTTCCAC 773
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATGAACTTCCAC 660

Qy        774 CTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAAATCA 833
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAAATCA 720

Qy        834 TGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCAATTGATATTCCAA 893
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCAATTGATATTCCAA 780

Qy        894 GTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAACAGCA 953
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAACAGCA 840

Qy        954 TCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGATTTCA 1013
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGATTTCA 900

Qy       1014 TGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAAGAAN 1073
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAAGAAC 960

Qy       1074 TAATAGGTTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGAGTGGATGATATTG 1133
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        961 TAATAGGTTTAGTTGTTCTTACCAAGTATAACAATAAGACATACAGAGTGGATGATATTG 1020

Qy       1134 ACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGGGGTCAGCTTCT 1193
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
```

```
Db      1021 ACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGAAGTCAGCTTCT 1080
Qy      1194 TAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTCTTGG 1253
Db      1081 TAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTCTTGG 1140
Qy      1254 TCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATGCTCA 1313
Db      1141 TCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATGCTCA 1200
Qy      1314 TTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAACGTGA 1373
Db      1201 TTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAACGTGA 1260
Qy      1374 TGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTGGGAC 1433
Db      1261 TGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTGGGAC 1320
Qy      1434 GACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGGGTT 1493
Db      1321 GACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGGGTT 1380
Qy      1494 TGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAGATTC 1553
Db      1381 TGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAGATTC 1440
Qy      1554 ACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAACAA 1613
Db      1441 ACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAACAA 1500
Qy      1614 GAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACGCGAA 1673
Db      1501 GAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACGCGAA 1560
Qy      1674 GAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGG 1733
Db      1561 GAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGG 1620
Qy      1734 GCATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACTGAAGCCTACTTAA 1793
Db      1621 GCATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACTGAAGCCTACTTAA 1680
Qy      1794 GAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCTGTTGTCAAGTA 1853
Db      1681 GAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCTGTTGTCAAGTA 1740
Qy      1854 ATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAA 1913
Db      1741 ATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAA 1800
Qy      1914 GTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAA 1973
Db      1801 GTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAA 1860
Qy      1974 AGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCC 2033
Db      1861 AGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCC 1920
Qy      2034 TGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGT 2093
Db      1921 TGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGT 1980
Qy      2094 CAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCA 2153
Db      1981 CAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCA 2040
Qy      2154 TATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGGCTC 2213
Db      2041 TATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGGCTC 2100
Qy      2214 TGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATCGTGTACCGCGATG 2273
Db      2101 TGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATCGTGTACCGCGATG 2160

Qy      2274 GCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCACAGTTTTTGGATT 2333
Db      2161 GCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCACAGTTTTTGGATT 2220
Qy      2334 GTCTAAAATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATTGTGGTGAAGAAAA 2393
Db      2221 GTCTAAAATCCATTGGTAGAGGTTACAACCCTAGACTAACGGTAATTGTGGTGAAGAAAA 2280
Qy      2394 GAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTCCTGGAA 2453
Db      2281 GAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTCCTGGAA 2340
Qy      2454 CAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTTATCGTGAGCCAGG 2513
Db      2341 CAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTTATCGTGAGCCAGG 2400
Qy      2514 CTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTATGACAACAGCGGCC 2573
Db      2401 CTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTATGACAACAGCGGCC 2460
Qy      2574 TGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATCTATTACAACTGGC 2633
Db      2461 TGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATCTATTACAACTGGC 2520
Qy      2634 CAG 2636
Db      2521 CAG 2523
```

```
RESULT 5
AF264004
LOCUS       AF264004               2328 bp    mRNA    linear   PRI 21-AUG-2001
DEFINITION  Homo sapiens HIWI mRNA, complete cds.
ACCESSION   AF264004
VERSION     AF264004.1  GI:15216446
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2328)
  AUTHORS   Sharma,A.K., Nelson,M.C., Brandt,J.E., Wessman,M., Muhmud,N.,
            Weller,K.P. and Hoffman,R.
  TITLE     Human CD34+ Stem Cells Express the hiwi Gene, a Human Homolog of
            the Drosophila Gene piwi
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2328)
  AUTHORS   Sharma,A.K. and Hoffman,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-2000) Medicine, UIC, 900 South Ashland Avenue,
            Chicago, IL 60607, USA
FEATURES             Location/Qualifiers
     source          1. .2328
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="testis"
     CDS             1. .2328
                     /codon_start=1
                     /product="HIWI"
                     /protein_id="AAK92281.1"
                     /db_xref="GI:15216447"
                     /translation="MIFGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALY
                     QYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTEVFSKTRNG
                     EDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRL
                     VIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKELI
                     GLVVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQPVL
                     VSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQRE
                     VGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADW
                     SKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDD
                     RTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQ
                     QTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASIN
                     EGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLK
```

TLVNYEVPQFLDCLKSIGRGYNPRLTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVE
VTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVI
RVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL"

BASE COUNT 716 a 496 c 520 g 596 t
ORIGIN

Query Match 66.6%; Score 2311; DB 9; Length 2328;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2314; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy       432 TTGGTGTGAATACAAGGCAGAACCTAGACCATGTTAAAGAATCAAAAACAGGTTCTTCAG 491
            ||||||||||| |||||||||||| ||||||||||||||||||||||||||||||||||||
Db         8 TTGGTGTGAACACAAGGCAGAACTTAGACCATGTTAAAGAATCAAAAACAGGTTCTTCAG 67

Qy       492 GCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGACATCCCGTCCCCAGTGGGCCT 551
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        68 GCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGACATCCCGTCCCCAGTGGGCCT 127

Qy       552 TATATCAGTATCACATTGACTATAACCCACTGATGGAAGCCAGAAGACTCCGTTCAGCTC 611
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       128 TATATCAGTATCACATTGACTATAACCCACTGATGGAAGCCAGAAGACTCCGTTCAGCTC 187

Qy       612 TTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCATGCTTTTGATGGAACGATATTAT 671
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       188 TTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCATGCTTTTGATGGAACGATATTAT 247

Qy       672 TTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTTTTTAGTAAGACCCGGAATGGAG 731
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       248 TTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTTTTTAGTAAGACCCGGAATGGAG 307

Qy       732 AGGATGTGAGGATAACGATCACTTTAACAAATGAACTTCCACCTACATCACCAACTTGTT 791
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       308 AGGATGTGAGGATAACGATCACTTTAACAAATGAACTTCCACCTACATCACCAACTTGTT 367

Qy       792 TGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAAATCATGAATTTGCAACAAATTG 851
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       368 TGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAAATCATGAATTTGCAACAAATTG 427

Qy       852 GACGAAATTATTATAACCCAAATGACCCAATTGATATTCCAAGTCACAGGTTGGTGATTT 911
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       428 GACGAAATTATTATAACCCAAATGACCCAATTGATATTCCAAGTCACAGGTTGGTGATTT 487

Qy       912 GGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAACAGCATCATGCTCTGCACTGACG 971
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       488 GGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAACAGCATCATGCTCTGCACTGACG 547

Qy       972 TTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGATTTCATGTTCAACTTTTATCATC 1031
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       548 TTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGATTTCATGTTCAACTTTTATCATC 607

Qy      1032 AGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAAGAANTAATAGGTTTAGTTGTTC 1091
            |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       608 AGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAAGAACTAATAGGTTTAGTTGTTC 667

Qy      1092 TTACCAAGTATAACATTAAGACATACAGAGTGGATGATATTGACTGGGACCAGAATCCCA 1151
            |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       668 TTACCAAGTATAACAATAAGACATACAGAGTGGATGATATTGACTGGGACCAGAATCCCA 727

Qy      1152 AGAGCACCTTTAAGAAAGCCGACGGCTCTGGGGTCAGCTTCTTAGAATACTACAGGAAGC 1211
            |||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db       728 AGAGCACCTTTAAGAAAGCCGACGGCTCTGAAGTCAGCTTCTTAGAATACTACAGGAAGC 787

Qy      1212 AATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGCCAGCCCAAGAGAA 1271
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       788 AATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGCCAGCCCAAGAGAA 847

Qy      1272 GGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATGCTCATTCCTGAGCTCTGCTATC 1331
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       848 GGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATGCTCATTCCTGAGCTCTGCTATC 907

Qy      1332 TTACAGGTCTAACTGATAAAATGCGTAATGATTTTAACGTGATGAAAGACTTAGCCGTTC 1391
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       908 TTACAGGTCTAACTGATAAAATGCGTAATGATTTTAACGTGATGAAAGACTTAGCCGTTC 967

Qy      1392 ATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTGGGACGACTCATTGATTACATTC 1451
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       968 ATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTGGGACGACTCATTGATTACATTC 1027

Qy      1452 ATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGGGTTTGAGCTTTGATTCCAACT 1511
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1028 ATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGGGTTTGAGCTTTGATTCCAACT 1087

Qy      1512 TACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAGATTCACCAAGGTGGAAAAACAT 1571
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1088 TACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAGATTCACCAAGGTGGAAAAACAT 1147

Qy      1572 TTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAACAAGAGGTGCACCATTAATTA 1631
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1148 TTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAACAAGAGGTGCACCATTAATTA 1207

Qy      1632 GTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACGCGAAGAAATTATGAAGCAGCCA 1691
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1208 GTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACGCGAAGAAATTATGAAGCAGCCA 1267

Qy      1692 ATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGCATGCAAATGAGAAAAG 1751
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1268 ATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGCATGCAAATGAGAAAAG 1327

Qy      1752 CAATAATGATTGAAGTGGATGACAGAACTGAAGCCTACTTAAGAGTCTTACAGCAAAAGG 1811
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1328 CAATAATGATTGAAGTGGATGACAGAACTGAAGCCTACTTAAGAGTCTTACAGCAAAAGG 1387

Qy      1812 TCACAGCAGACACCCAGATAGTTGTCTGTCTGTTGTCAAGTAATCGGAAGGACAAATACG 1871
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1388 TCACAGCAGACACCCAGATAGTTGTCTGTCTGTTGTCAAGTAATCGGAAGGACAAATACG 1447

Qy      1872 ATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCC 1931
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1448 ATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCC 1507

Qy      1932 GAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAAAGATTGCCCTACAGATGA 1991
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1508 GAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAAAGATTGCCCTACAGATGA 1567

Qy      1992 ACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCTCGTGATGATCG 2051
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1568 ACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCTCGTGATGATCG 1627

Qy      2052 TTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTG 2111
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1628 TTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTG 1687

Qy      2112 CCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCATATTTCAGGATAGAGGAC 2171
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1688 CCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCATATTTCAGGATAGAGGAC 1747

Qy      2172 AGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGGCTCTGAGGGCTTGGAATAGCT 2231
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1748 AGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGGCTCTGAGGGCTTGGAATAGCT 1807

Qy      2232 GCAATGAGTACATGCCCAGCCGGATCATCGTGTACCGCGATGGCGTAGGAGACGGCCAGC 2291
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1808 GCAATGAGTACATGCCCAGCCGGATCATCGTGTACCGCGATGGCGTAGGAGACGGCCAGC 1867

Qy      2292 TGAAAACACTGGTGAACTACGAAGTGCCACAGTTTTTGGATTGTCTAAAATCCATTGGTA 2351
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1868 TGAAAACACTGGTGAACTACGAAGTGCCACAGTTTTTGGATTGTCTAAAATCCATTGGTA 1927

Qy      2352 GAGGTTACAACCCTAGANTAACGGTAATTGTGGTGAAGAAAAGAGTGAACACCAGATTTT 2411
            |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      1928 GAGGTTACAACCCTAGACTAACGGTAATTGTGGTGAAGAAAAGAGTGAACACCAGATTTT 1987

Qy      2412 TTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTCCTGGAACAGTTATTGATGTAGAGG 2471
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1988 TTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTCCTGGAACAGTTATTGATGTAGAGG 2047
```

```
Qy       2472 TTACCAGACCAGAATGGTATGACTTTTTTATCGTGAGCCAGGCTGTGAGAAGTGGTAGTG 2531
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2048 TTACCAGACCAGAATGGTATGACTTTTTTATCGTGAGCCAGGCTGTGAGAAGTGGTAGTG 2107

Qy       2532 TTTCTCCCACACATTACAATGTCATCTATGACAACAGCGGCCTGAAGCCAGACCACATAC 2591
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2108 TTTCTCCCACACATTACAATGTCATCTATGACAACAGCGGCCTGAAGCCAGACCACATAC 2167

Qy       2592 AGCGCTTGACCTACAAGCTGTGCCACATCTATTACAACTGGCCAGGTGTCATTCGTGTTC 2651
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2168 AGCGCTTGACCTACAAGCTGTGCCACATCTATTACAACTGGCCAGGTGTCATTCGTGTTC 2227

Qy       2652 CTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTGTTGGCCAGAGTATTCACAGAG 2711
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2228 CTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTGTTGGCCAGAGTATTCACAGAG 2287

Qy       2712 AGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAA 2752
             ||||||||||||||||||||||||||||||||||||||||
Db       2288 AGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAA 2328
```

```
RESULT 6
AB032604
LOCUS       AB032604                3910 bp    mRNA    linear   ROD 04-OCT-2001
DEFINITION  Mus musculus mRNA for MIWI, complete cds.
ACCESSION   AB032604
VERSION     AB032604.1  GI:7416110
KEYWORDS    MIWI (piwi).
SOURCE      Mus musculus tissue_lib:testis cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Kuramochi-Miyagawa,S., Kimura,T., Yomogida,K., Kuroiwa,A.,
            Tadokoro,Y., Fujita,Y., Sato,M., Matsuda,Y. and Nakano,T.
  TITLE     Two mouse piwi-related genes: miwi and mili
  JOURNAL   Mech. Dev. 108 (1-2), 121-133 (2001)
  MEDLINE   21463379
REFERENCE   2  (bases 1 to 3910)
  AUTHORS   Miyagawa,S.K. and Nakano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-1999) Satomi K Miyagawa, Research Institute for
            Microbial Diseases, Department of Molecular Cell Biology; Yamadaoka
            3-1, Suita, Osaka 565-0871, Japan
            (E-mail:smiyagaw@biken.osaka-u.ac.jp, Tel:81-6-6879-8362,
            Fax:81-6-6879-8362)
FEATURES             Location/Qualifiers
     source          1. .3910
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /map="5F distal-G2 proximal"
                     /tissue_lib="testis"
     gene            1. .3910
                     /gene="miwi"
     CDS             156. .2744
                     /gene="miwi"
                     /note="piwi family"
                     /codon_start=1
                     /product="MIWI"
                     /protein_id="BAA93705.1"
                     /db_xref="GI:7416111"
                     /translation="MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEG
                     DLVGRGRQRGMVVGATSKSQELQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVK
                     ESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGR
                     CHAFDGTILFLPKRLQHKVTEVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIIFR
                     RLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLR
                     SETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFK
                     KADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTG
                     LTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNL
                     LSFSGRILQSEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEA
                     ANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTEAYLRALQQKVTSDTQIVVCLLSSNRK
                     DKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMPL
                     KLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQA
                     ALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIV
                     VKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVI
                     YDSSGLKPDHIQRLTYKLCHVYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLS
                     NRLYYL"
BASE COUNT      993 a    970 c   1063 g    884 t
ORIGIN
```

```
  Query Match             53.0%;  Score 1839.2;  DB 10;  Length 3910;
  Best Local Similarity   82.0%;  Pred. No. 0;
  Matches 2130;  Conservative    0;  Mismatches  466;  Indels    3;  Gaps    1;

Qy        164 ACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCG 223
              | |||||||||  ||||||||||  | ||| | || ||||| || ||||||||||| | |
Db        153 AAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGAGACGGTG 212

Qy        224 CAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCG 283
              ||||  || ||  |  |||| || |||||||||| || || |  || || ||||| | |
Db        213 CAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACCTCAACAG 272

Qy        284 CCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAA---CAGCAGGAGGA 340
              | ||  |||||||||| ||  ||||||| ||||| ||||| || |    ||  ||||
Db        273 TCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGTCGGAGCC 332

Qy        341 ACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAG 400
              ||| |||||||||||| ||| ||||| || ||||| |||||||| | ||  | ||||||
Db        333 ACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACTGGCAGAG 392

Qy        401 AGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGAC 460
              ||||||||  |||||  |||| || ||||| ||||||||||| || || |||||||| |||
Db        393 AGAGGAGGGCGTCGCCGAGACTTCCATGACCTTGGTGTGAACACCAGACAGAACCTTGAC 452

Qy        461 CATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCAT 520
              ||||| ||||| ||||| ||||| || || ||||| || || | | | ||||| |||||
Db        453 CATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCAC 512

Qy        521 TTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCA 580
              |||||||||| || || || |||||||| | |||||||| ||||| ||||| || ||
Db        513 TTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCC 572

Qy        581 CTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGA 640
              |||||||| ||| |||| || || || || || || || || || ||||| || ||||||
Db        573 CTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGA 632

Qy        641 AAGTGTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTT 700
              | |||||||||||| ||||| || ||||| |||||||||| |||||||||| |||||
Db        633 AGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGTC 692

Qy        701 ACTGAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACA 760
              || ||||| || ||| |||| |||||||| ||  | |||||||| || ||||  | ||
Db        693 ACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCACCCTGACC 752

Qy        761 AATGAACTTCCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGG 820
              || || || || || || || || || ||  |||||||||||| || || ||||||||
Db        753 AACGAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCATCTTCAGGAGG 812

Qy        821 CTTTTGAAAATCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCA 880
              || |||||||||||||||||||||||||||||||| |||||||| || |||| ||||||
Db        813 CTCTTGAAAATCATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAGTGACCCG 872

Qy        881 ATTGATATTCCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAG 940
              |||||||||||||  |||||||||||||| ||||| |||||||| || ||||||||||||
Db        873 ATTGATATTCCAAACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCATCCTTCAG 932

Qy        941 TATGAAAACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACT 1000
              ||||| |||| ||||||||||||||| ||||| ||||| || || || || || ||||||
Db        933 TATGAGAACAACATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAGCGAGACT 992

Qy       1001 GTTTTGGATTTCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAA 1060
              ||  | || |||||||||||  | || || |||||||| || || || || || || |||
Db        993 GTCCTAGACTTCATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCAGGAGCAA 1052

Qy       1061 GTTTCCAAAGAANTAATAGGTTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGA 1120
              || || || ||  | |||||  |  | ||||| ||||||||| || |  ||||| ||| |
```

```
Db       1053 GTGTCGAAGGAGCTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGACCTACCGG 1112

Qy       1121 GTGGATGATATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCT 1180
             ||||||||| |||||||||||||||||||| |||||||||| ||||| || || |||||
Db       1113 GTGGATGACATTGACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGATGGCTCG 1172

Qy       1181 GGGGTCAGCTTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAG 1240
             | |||||||||| | || |||||||||||||||||||||||| |||||||| ||| |||||
Db       1173 GAGGTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGACCTGAAG 1232

Qy       1241 CAGCCTGTCTTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGG 1300
             ||||| ||  |||| ||||| |||||| | ||| | ||||| || || || ||||| ||
Db       1233 CAGCCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCTGCCTGGC 1292

Qy       1301 CCTGCCATGCTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAAT 1360
             || || |||||||| |||| |||||||||| |||| || |||||||||||||| ||| |||
Db       1293 CCAGCTATGCTCATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAATGCGCAAT 1352

Qy       1361 GATTTTAACGTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAG 1420
             ||||| || |||||||| ||| | || || || ||  | || || || ||||  ||||
Db       1353 GATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCAGCGGCAG 1412

Qy       1421 CGTGAAGTGGGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTT 1480
             || || ||||| || |||| || ||||| || ||  | || |||| || || ||||||
Db       1413 CGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAGAGAGCTT 1472

Qy       1481 CGAGACTGGGGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAA 1540
             ||||||||||  ||||||| || || |||| |||||||||| |||||||| || |||
Db       1473 CGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAATCTTACAA 1532

Qy       1541 ACAGAAAAGATTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGG 1600
             | || |||| |||| || |||| || |||||||||| |||||||||||| |||
Db       1533 TCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGCAGACTGG 1592

Qy       1601 TCCAAAGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTG 1660
             |||||||||||||||| || ||  | || || || |||||| | |||||||||||| ||
Db       1593 TCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGGCTGCTG 1652

Qy       1661 ATCTATACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTT 1720
             |||||||| || || |||||||||||||||||| ||| ||||||| || || || |||||
Db       1653 ATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTTCAAAGTG 1712

Qy       1721 ACACCAGCCATGGGCATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACT 1780
             || |||||||||||||| || |||| ||| ||||| ||||| || ||||||||||||||
Db       1713 ACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGACAGAACA 1772

Qy       1781 GAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGT 1840
             ||||| ||  | |||| ||| ||||| ||||| ||  ||||||| ||||||||||||||
Db       1773 GAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGT 1832

Qy       1841 CTGTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGAT 1900
             || |||||||||||||||||||||||||||| ||||| || || || ||| ||||||||||
Db       1833 CTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTGTACAGAC 1892

Qy       1901 TGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATG 1960
             ||||| |||||||||||||||||||||||||| ||| | ||||| |||||||| ||||||
Db       1893 TGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATG 1952

Qy       1961 GCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGG 2020
             |||||||| || ||||| ||||| |||||||||||||||||||||| ||||||||| ||
Db       1953 GCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGG 2012

Qy       2021 GTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACA 2080
             |||||||| |||||||| || |  |||||||| |||||||| ||||||||||||  |||
Db       2013 GTGGACATGCCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACACCACA 2072

Qy       2081 GCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGG 2140
             |||||||||||||| |||||||||| || ||||||||||||||||||||||||||||||
Db       2073 GCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGG 2132

Qy       2141 TTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGC 2200
             ||||| |||||| | |||||||| | || |||||||||| ||||| ||||| || |||
Db       2133 TTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGC 2192
```

```
Qy       2201 CTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATC 2260
             ||||||| |||||||||||||||| | |||||||||| |||||||||||||  ||||
Db       2193 TTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATC 2252

Qy       2261 GTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCA 2320
             |||||||| || || || || ||||| |||||||| || ||||| || || || || |||
Db       2253 GTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCA 2312

Qy       2321 CAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATT 2380
             |||||  | ||||| || || ||  | || |||||||||||||| ||| | || |||||
Db       2313 CAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATC 2372

Qy       2381 GTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAAT 2440
             |||||||||||  | || ||  |||| |||||||||||||||| ||||||||||||||
Db       2373 GTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAAC 2432

Qy       2441 CCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTT 2500
             || ||||| || ||||| || ||||| || || |||||||||| ||||||||||||||
Db       2433 CCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTC 2492

Qy       2501 ATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTAT 2560
             |||||||||||||| |||||||| || |||| || || |||| ||||||||||||||||
Db       2493 ATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTAT 2552

Qy       2561 GACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATC 2620
             |||| ||| |||||||||| |||||||| |||||  |||| |||||||| |||||| |
Db       2553 GACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGCTCTGCCACGTG 2612

Qy       2621 TATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTG 2680
             || || || |||| || |||| || || |||| |||||||||| || |||||||||
Db       2613 TACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTG 2672

Qy       2681 GCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTT 2740
             || || || || |||||||| || |||||||||||||| || || ||||| ||||||||
Db       2673 GCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTC 2732

Qy       2741 TACTACCTCTAACCTGCAG 2759
             |||||||||||||| ||||
Db       2733 TACTACCTCTAACCAGCAG 2751
```

```
RESULT 7
AF438405
LOCUS       AF438405               4064 bp    mRNA    linear   ROD 13-NOV-2001
DEFINITION  Mus musculus MIWI mRNA, complete cds.
ACCESSION   AF438405
VERSION     AF438405.1  GI:16905060
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4064)
  AUTHORS   Deng,W. and Lin,H.
  TITLE     Miwi, a murine homolog of piwi, encodes a cytoplasmic protein
            essential for spermatogenesis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4064)
  AUTHORS   Deng,W. and Lin,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2001) Department of Cell Biology, Duke
            University, Room 412, Nanaline Duke Building, Durham, NC 27710, USA
FEATURES             Location/Qualifiers
     source          1. .4064
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /tissue_type="testis"
     CDS             191. .2779
                     /note="Piwi-like protein"
                     /codon_start=1
                     /product="MIWI"
```

/protein_id="AAL31014.1"
/db_xref="GI:16905061"
/translation="MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEG
DLVGRGRQRGMVVGATSKSQELQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVK
ESKTGSSGIIVKLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGR
CHAFDGTILFLPKRLQHKVTEVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIIFR
RLLKIMNLQQIGRNYYNPSDPIDIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLR
SETVLDFMFNLYQQTEEHKFQEQVSKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFK
KADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTG
LTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKDDNVQRELRDWGLSFDSNL
LSFSGRILQSEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEA
ANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTEAYLRALQQKVTSDTQIVVCLLSSNRK
DKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDMPL
KLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWFSRCVFQDRGQELVDGLKVCLQA
ALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQFLDCLKSVGRGYNPRLTVIV
VKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVI
YDSSGLKPDHIQRLTYKLCHVYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLS
NRLYYL"
BASE COUNT 1115 a 982 c 1077 g 890 t
ORIGIN

Query Match 53.0%; Score 1839.2; DB 10; Length 4064;
Best Local Similarity 82.0%; Pred. No. 0;
Matches 2130; Conservative 0; Mismatches 466; Indels 3; Gaps 1;

```
Qy        164 ACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCG 223
              | |||||||||  ||||||||||  | ||| | || ||||| || ||||||||||| | |
Db        188 AAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGAGACGGTG 247

Qy        224 CAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCG 283
              ||||  || ||  |  |||| || |||||||||||| || || |  || || ||||| | |
Db        248 CAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACCTCAACAG 307

Qy        284 CCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAA---CAGCAGGAGGA 340
              | ||  |||||||||| ||  ||||||| ||||| ||||| || |    ||  ||||
Db        308 TCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGTCGGAGCC 367

Qy        341 ACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAG 400
              ||| |||||||||||| ||| ||||| || ||||| ||||||||| | ||  | ||||||
Db        368 ACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACTGGCAGAG 427

Qy        401 AGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGAC 460
              ||||||||| |||||  |||| || ||||| |||||||||| || || |||||||| |||
Db        428 AGAGGAGGGCGTCGCCGAGACTTCCATGACCTTGGTGTGAACACCAGACAGAACCTTGAC 487

Qy        461 CATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCAT 520
              ||||| ||||| ||||| ||||| || || ||||| || || | | | ||||| |||||
Db        488 CATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCAC 547

Qy        521 TTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCA 580
              |||||||||| || || || ||||||||| | |||||||| ||||| ||||| || ||
Db        548 TTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCC 607

Qy        581 CTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGA 640
              |||||||| ||| |||| || || || || || || || || || ||||| || ||||||
Db        608 CTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGA 667

Qy        641 AAGTGTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTT 700
              | |||||||||||| ||||| || ||||| |||||||||| |||||||||| |||||
Db        668 AGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGTC 727

Qy        701 ACTGAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACA 760
              || ||||| || ||| |||| |||||||| ||  | |||||||| || |||||  | ||
Db        728 ACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCACCCTGACC 787

Qy        761 AATGAACTTCCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGG 820
              || || || || || || || || || ||  |||||||||||| || || |||||||||
Db        788 AACGAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCATCTTCAGGAGG 847

Qy        821 CTTTTGAAAATCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCA 880
              || ||||||||||||||||||||||||||||||||| |||||||| || |||| ||||||
Db        848 CTCTTGAAAATCATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAGTGACCCG 907

Qy        881 ATTGATATTCCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAG 940
              |||||||||||||  |||||||||||||| ||||| |||||||| || ||||||||||||
Db        908 ATTGATATTCCAAACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCATCCTTCAG 967

Qy        941 TATGAAAACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACT 1000
              ||||| |||| |||||||||||||| ||||| ||||| || || || || || ||||||
Db        968 TATGAGAACAACATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAGCGAGACT 1027

Qy       1001 GTTTTGGATTTCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAA 1060
              ||  | || ||||||||||  | || || |||||||| || || || || || || |||
Db       1028 GTCCTAGACTTCATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCAGGAGCAA 1087

Qy       1061 GTTTCCAAAGAANTAATAGGTTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGA 1120
              || || || ||  | |||||  |  | ||||| |||||||| || |  ||||| ||| |
Db       1088 GTGTCGAAGGAGCTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGACCTACCGG 1147

Qy       1121 GTGGATGATATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCT 1180
              |||||||| |||||||||||||||||||| |||||||||| ||||| || || |||||
Db       1148 GTGGATGACATTGACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGATGGCTCG 1207

Qy       1181 GGGGTCAGCTTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAG 1240
              | |||||||||| | || |||||||||||||||||||||| |||||||| ||| |||||
Db       1208 GAGGTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGACCTGAAG 1267

Qy       1241 CAGCCTGTCTTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGG 1300
              ||||| ||  |||| |||||  |||||| | ||| | |||||  || || || ||||| ||
Db       1268 CAGCCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCTGCCTGGC 1327

Qy       1301 CCTGCCATGCTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAAT 1360
              || || |||||||| ||||| |||||||||| ||||| || |||||||||||||| |||
Db       1328 CCAGCTATGCTCATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAATGCGCAAT 1387

Qy       1361 GATTTTAACGTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAG 1420
              ||||| || |||||||| ||| | || || || ||  | || || || |||||  |||||
Db       1388 GATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCAGCGGCAG 1447

Qy       1421 CGTGAAGTGGGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTT 1480
              || || ||||| || ||||| || ||||| || ||  | || ||||| || || ||||||
Db       1448 CGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAGAGAGCTT 1507

Qy       1481 CGAGACTGGGGGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAA 1540
              ||||||||||  |||||| || || ||||| |||||||||| |||||||| || |||
Db       1508 CGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAATCTTACAA 1567

Qy       1541 ACAGAAAAGATTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGG 1600
              | || ||||| ||||| || ||||| || |||||||||| |||||||||||||| |||
Db       1568 TCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGCAGACTGG 1627

Qy       1601 TCCAAAGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTG 1660
              |||||||||||||||||| || ||  || || || |||||| | ||||||||||| ||
Db       1628 TCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGGCTGCTG 1687

Qy       1661 ATCTATACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTT 1720
              |||||||| || || ||||||||||||||||| ||| ||||||| || || || |||||
Db       1688 ATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTTCAAAGTG 1747

Qy       1721 ACACCAGCCATGGGCATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACT 1780
              || |||||||||||||| || |||| ||| ||||| ||||| || ||||||||||||||
Db       1748 ACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGACAGAACA 1807

Qy       1781 GAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGT 1840
              ||||| ||  | |||| ||| ||||| ||||| ||  ||||||| ||||||||||||||
Db       1808 GAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGT 1867

Qy       1841 CTGTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGAT 1900
              || |||||||||||||||||||||||||||| ||||| || || || ||| ||||||||||
Db       1868 CTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTGTACAGAC 1927

Qy       1901 TGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATG 1960
              ||||| |||||||||||||||||||||||||||| ||| | ||||| ||||||||| |||||
Db       1928 TGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATG 1987

Qy       1961 GCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGG 2020
```

```
Db       1988 GCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGG 2047

Qy       2021 GTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACA 2080

Db       2048 GTGGACATGCCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACACCACA 2107

Qy       2081 GCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGG 2140

Db       2108 GCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGG 2167

Qy       2141 TTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGC 2200

Db       2168 TTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGC 2227

Qy       2201 CTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATC 2260

Db       2228 TTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATC 2287

Qy       2261 GTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCA 2320

Db       2288 GTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCA 2347

Qy       2321 CAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATT 2380

Db       2348 CAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATC 2407

Qy       2381 GTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAAT 2440

Db       2408 GTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAAC 2467

Qy       2441 CCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTT 2500

Db       2468 CCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTC 2527

Qy       2501 ATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTAT 2560

Db       2528 ATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTAT 2587

Qy       2561 GACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATC 2620

Db       2588 GACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGCTCTGCCACGTG 2647

Qy       2621 TATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTG 2680

Db       2648 TACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTG 2707

Qy       2681 GCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTT 2740

Db       2708 GCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTC 2767

Qy       2741 TACTACCTCTAACCTGCAG 2759

Db       2768 TACTACCTCTAACCAGCAG 2786


RESULT 8
AF336369
LOCUS       AF336369               3273 bp    mRNA    linear   VRT 02-JAN-2002
DEFINITION  Danio rerio piwi protein mRNA, complete cds.
ACCESSION   AF336369
VERSION     AF336369.1  GI:18033486
KEYWORDS    .
SOURCE      Danio rerio.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 3273)
  AUTHORS   Weeraratne,S.D., Gong,Z. and Tan,C.-H.
  TITLE     Cloning and characterization of zebrafish homolog of piwi,
            essential for germ-line stem cell self-renewal
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3273)
  AUTHORS   Weeraratne,S.D., Gong,Z. and Tan,C.-H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2001) Department of Biological Sciences, National
            University of Singapore, 10 Kent Ridge Crescent, Singapore 119260,
            Singapore
FEATURES             Location/Qualifiers
     source          1. .3273
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
     CDS             64. .2640
                     /codon_start=1
                     /product="piwi protein"
                     /protein_id="AAL57170.1"
                     /db_xref="GI:18033487"
                     /translation="MTGRARARSRGRGRGQEPAAPGAQPPVSQEAAKPVVSTPSEGQL
                     VGRGRQKPAPGAMSEEAMLQISAGFQQVKIGERGGRRRDFHDSGVHTRQLMEHVKESK
                     TGVSGTAIELRANFMRLLSRPMWALYQYHVDYKPPMESRRLRSALLFQHEETLGKAHT
                     FDGAILFLPNKLRNAETVLCSETRNGEKVEITVTLTNELPPSSPVCLQFYNILFRRIL
                     RILNMQQIGRHYYNPDDPFNIPQHRLTIWPGFMTTILQYESSIMLCSDVSHKVLRSET
                     VLDFMYSLRQQCGDQRFPEACTKELVGLIILTKYNNKTYRIDDIAWDHTPNNTFKKGD
                     TEISFKNYFKSQYGLDITDGNQVLLVSHVKRLGPSGRPPPGPAMLVPEFCYLTGLTDK
                     MRADFNIMKDLASHTRLSPEQREGRINRLISNINRNGDVQNELTTWGLSFENKLLSLN
                     GRVLPSERIIQGGRAFEYNPWTADWSKEMRGLPLISCMSLDNWLMFYTRRNADVAQSL
                     LQTLNKVSGPMGIRMQRAVMIEYEDRQESLLRALQQNVARETQMVVVILPTNRKDKYD
                     CVKKYLCVDCPTPSQCVVSRTISKPQALMTVATKIALQMNCKMGGELWSVEIPLRQLM
                     IVGIDCYHDTAAGKRSIGAMVASLNQGMSRWFSKCVLQNRGQEIIDALKGSLQGALKA
                     YLKYNNSLPSRIIVYRDGVGDGMLQSVVDYEVPQIMQSIKTMGQDYEPKLSVVVVKKR
                     ISSRFFARIDGKIANPPPGTVIDTEVTRPEWYDFFIVSQAVRFGCVAPTHYNVVFDNS
                     GLKPDHMQRLTYKLCHMYYNWQGIVRVPAPCQYAHKLAFLVGQSIHKEPNMNLDDFLY
                     YL"
BASE COUNT      922 a    717 c    787 g    847 t
ORIGIN

  Query Match             30.7%;  Score 1067.2;  DB 5;  Length 3273;
  Best Local Similarity   63.3%;  Pred. No. 6.6e-239;
  Matches 1652;  Conservative    0;  Mismatches  956;  Indels    3;  Gaps    1;

Qy       153 AGAAATAGAAAACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTC 212

Db        44 AGCATTTGTAAAGCACCGGTATGACAGGACGAGCAAGAGCAAGATCGAGGGGCAGAGGTC 103

Qy       213 AGGAGACAGCGCAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTA 272

Db       104 GTGGTCAGGAGCCTGCTGCTCCTGGCGCACAACCACCTGTATCTCAAGAGGCTGCAAAGC 163

Qy       273 GGCCTCAGCCGCCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAG 332

Db       164 CTGTTGTGTCCACTCCGTCTGAAGGACAACTGGTGGGCAGAGGGAGACAGAAACCTGCAC 223

Qy       333 CAGGAGGAACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGN 392

Db       224 CTGGAGCCATGTCTGAAGAAGCGATGCTCCAGATTTCTGCTGGCTTTCAGCAGGTGAAGA 283

Qy       393 TAGCAGAGAGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGA 452

Db       284 TCGGGGAGAGAGGTGGACGCAGACGGGATTTCCACGATTCTGGTGTCCATACTCGTCAAC 343

Qy       453 ACCTAGACCATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCA 512

Db       344 TTATGGAACACGTGAAAGAATCCAAAACAGGTGTTTCTGGGACTGCCATTGAGCTGAGGG 403

Qy       513 CTAACCATTTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACT 572

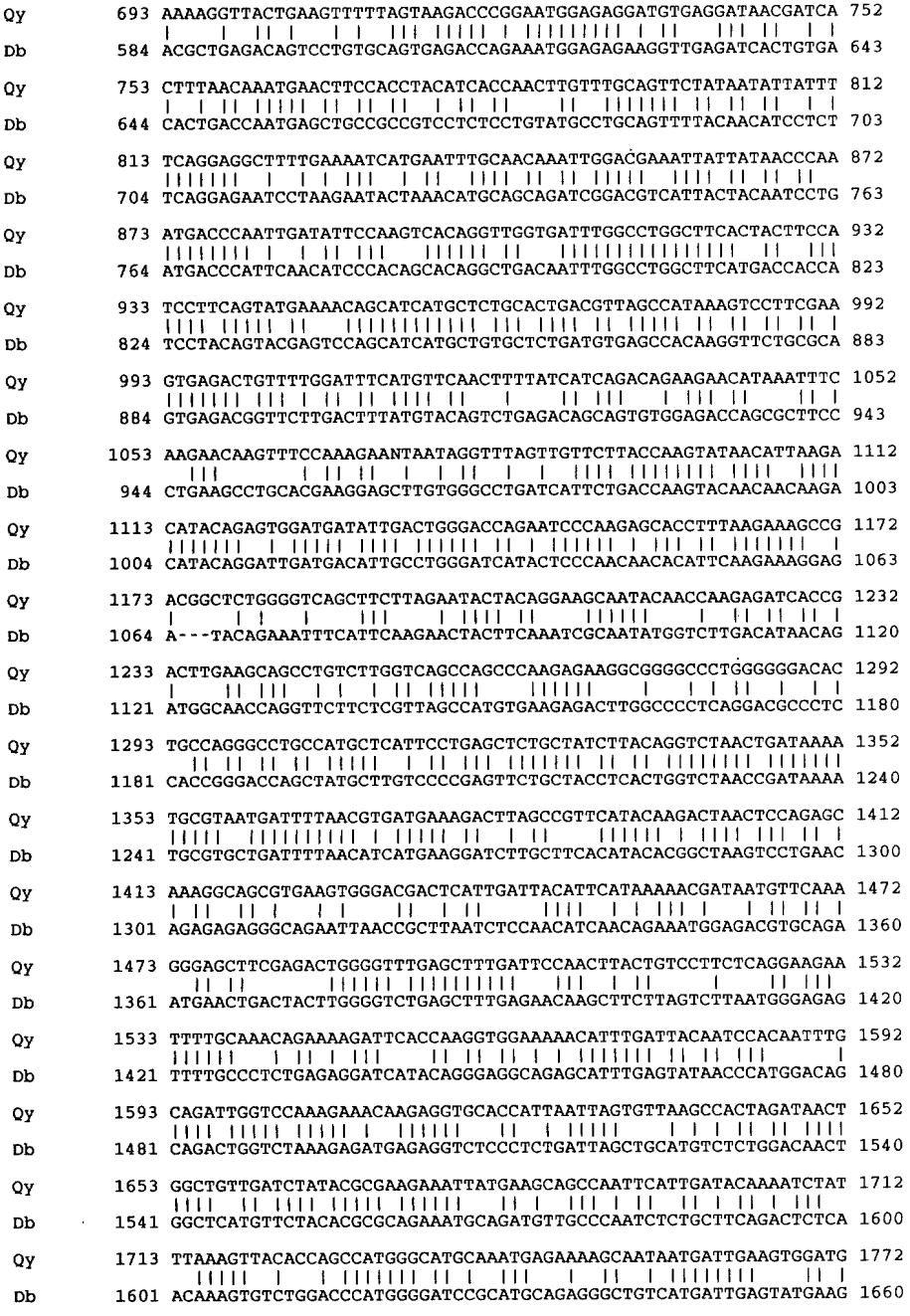
Db       404 CCAATTTTATGCGTCTGCTTTCACGTCCTATGTGGGCTTTGTATCAGTATCATGTGGACT 463

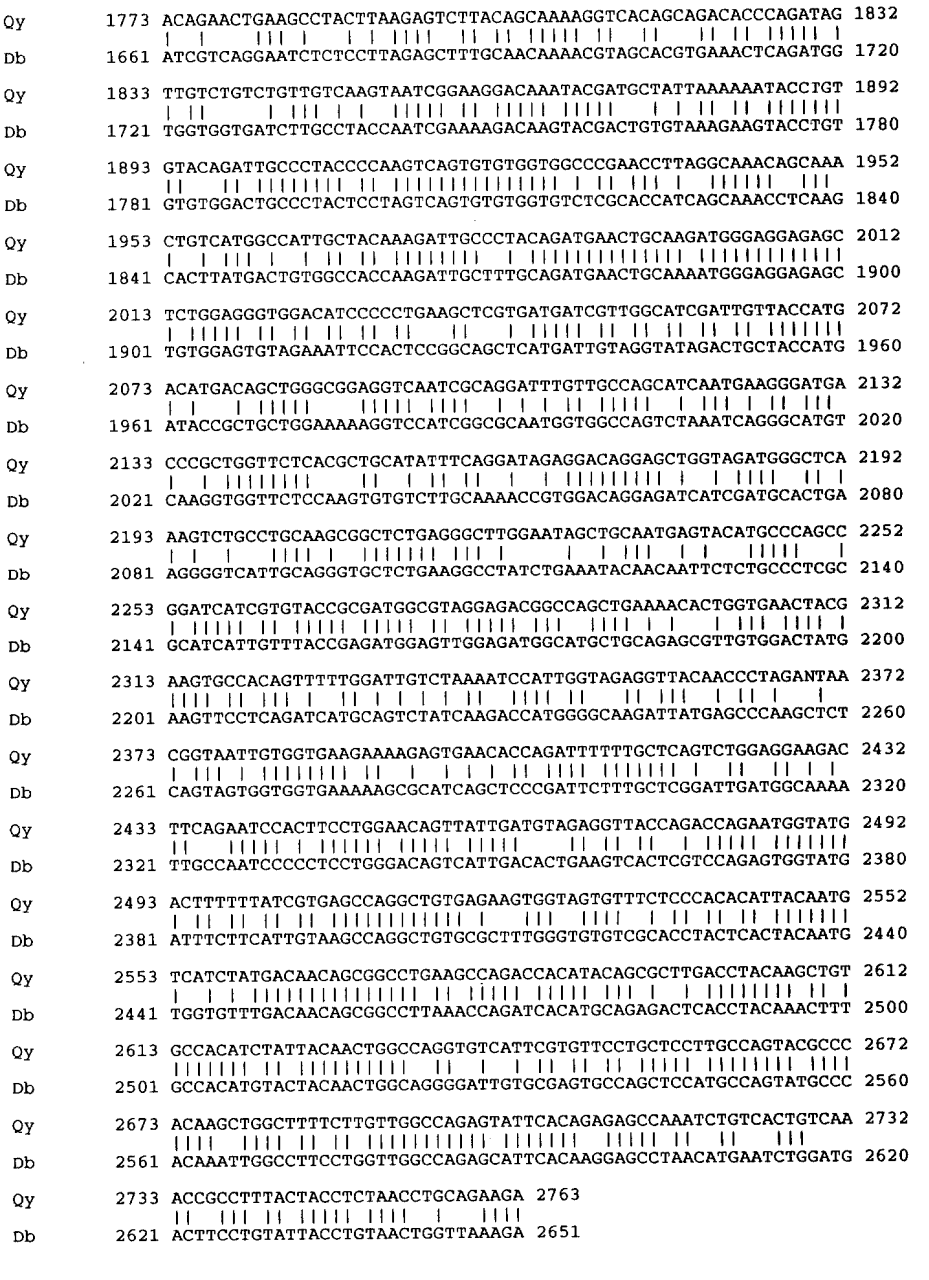
Qy       573 ATAACCCACTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATC 632

Db       464 ACAAGCCACCAATGGAGTCCAGACGACTGCGTTCGGCTTTGCTTTTCCAGCACGAGGAGA 523

Qy       633 TAATTGGAAAGTGTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGC 692

Db       524 CTCTGGGGAAAGCACACACCTTTGATGGAGCCATTCTCTTTCTCCCCAACAAACTGCGCA 583
```





RESULT 9
AC025837/c
LOCUS AC025837 193697 bp DNA linear HTG 06-MAY-2001
DEFINITION Homo sapiens clone RP11-200K12, WORKING DRAFT SEQUENCE, 29

```
            unordered pieces.
ACCESSION   AC025837
VERSION     AC025837.3  GI:8072609
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 193697)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome, clone RP11-200K12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 193697)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7342026.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5437
                Center clone name: 200_K_12
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 174594 bases at least Q40
                Consensus quality: 183716 bases at least Q30
                Consensus quality: 187939 bases at least Q20
                Insert size: 183000; agarose-fp
                Insert size: 190897; sum-of-contigs
                Quality coverage: 3.9 in Q20 bases; agarose-fp
                Quality coverage: 3.7 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 29 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1628: contig of 1628 bp in length
            *     1629 1728: gap of      100 bp
            *     1729     2962: contig of 1234 bp in length
            *     2963 3062: gap of      100 bp
            *     3063     4989: contig of 1927 bp in length
            *     4990 5089: gap of      100 bp
            *     5090     6666: contig of 1577 bp in length
            *     6667 6766: gap of      100 bp
            *     6767     8667: contig of 1901 bp in length
            *     8668 8767: gap of      100 bp
            *     8768    11259: contig of 2492 bp in length
            *    11260 11359: gap of      100 bp
            *    11360    13557: contig of 2198 bp in length
            *    13558 13657: gap of      100 bp
            *    13658    17173: contig of 3516 bp in length
            *    17174 17273: gap of      100 bp
            *    17274    20060: contig of 2787 bp in length
            *    20061 20160: gap of      100 bp
            *    20161    23843: contig of 3683 bp in length
            *    23844 23943: gap of      100 bp
            *    23944    27203: contig of 3260 bp in length
            *    27204 27303: gap of      100 bp
            *    27304    31592: contig of 4289 bp in length
            *    31593 31692: gap of      100 bp
            *    31693    36740: contig of 5048 bp in length
            *    36741 36840: gap of      100 bp
            *    36841    42197: contig of 5357 bp in length
            *    42198 42297: gap of      100 bp
            *    42298    47652: contig of 5355 bp in length
            *    47653 47752: gap of      100 bp
            *    47753    52307: contig of 4555 bp in length
            *    52308 52407: gap of      100 bp
            *    52408    58606: contig of 6199 bp in length
            *    58607 58706: gap of      100 bp
            *    58707    65584: contig of 6878 bp in length
            *    65585 65684: gap of      100 bp
            *    65685    72492: contig of 6808 bp in length
            *    72493 72592: gap of      100 bp
            *    72593    78756: contig of 6164 bp in length
            *    78757 78856: gap of      100 bp
            *    78857    85474: contig of 6618 bp in length
            *    85475 85574: gap of      100 bp
            *    85575    91951: contig of 6377 bp in length
            *    91952 92051: gap of      100 bp
            *    92052    99044: contig of 6993 bp in length
            *    99045 99144: gap of      100 bp
            *    99145   104884: contig of 5740 bp in length
            *   104885 104984: gap of      100 bp
            *   104985   112130: contig of 7146 bp in length
            *   112131 112230: gap of      100 bp
            *   112231   118017: contig of 5787 bp in length
            *   118018 118117: gap of      100 bp
            *   118118   126319: contig of 8202 bp in length
            *   126320 126419: gap of      100 bp
            *   126420   137651: contig of 11232 bp in length
            *   137652 137751: gap of      100 bp
            *   137752   193697: contig of 55946 bp in length.
FEATURES             Location/Qualifiers
     source          1. .193697
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-200K12"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1628
                     /note="assembly_fragment"
     misc_feature    1729. .2962
                     /note="assembly_fragment"
     misc_feature    3063. .4989
                     /note="assembly_fragment"
     misc_feature    5090. .6666
                     /note="assembly_fragment"
     misc_feature    6767. .8667
                     /note="assembly_fragment"
     misc_feature    8768. .11259
                     /note="assembly_fragment"
     misc_feature    11360. .13557
                     /note="assembly_fragment"
```

```
     misc_feature    13658. .17173
                     /note="assembly_fragment"
     misc_feature    17274. .20060
                     /note="assembly_fragment"
     misc_feature    20161. .23843
                     /note="assembly_fragment"
     misc_feature    23944. .27203
                     /note="assembly_fragment"
     misc_feature    27304. .31592
                     /note="assembly_fragment"
     misc_feature    31693. .36740
                     /note="assembly_fragment"
     misc_feature    36841. .42197
                     /note="assembly_fragment"
     misc_feature    42298. .47652
                     /note="assembly_fragment"
     misc_feature    47753. .52307
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    52408. .58606
                     /note="assembly_fragment"
     misc_feature    58707. .65584
                     /note="assembly_fragment"
     misc_feature    65685. .72492
                     /note="assembly_fragment"
     misc_feature    72593. .78756
                     /note="assembly_fragment"
     misc_feature    78857. .85474
                     /note="assembly_fragment"
     misc_feature    85575. .91951
                     /note="assembly_fragment"
     misc_feature    92052. .99044
                     /note="assembly_fragment"
     misc_feature    99145. .104884
                     /note="assembly_fragment"
     misc_feature    104985. .112130
                     /note="assembly_fragment"
     misc_feature    112231. .118017
                     /note="assembly_fragment"
     misc_feature    118118. .126319
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    126420. .137651
                     /note="assembly_fragment"
     misc_feature    137752. .193697
                     /note="assembly_fragment"
BASE COUNT    55529 a  42321 c  40366 g  52678 t   2803 others
ORIGIN

  Query Match             23.9%;  Score 830.8;  DB 2;  Length 193697;
  Best Local Similarity   99.8%;  Pred. No. 2.4e-183;
  Matches  832;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2636 GGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTGTTGGC 2695
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138962 GGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTGTTGGC 138903

Qy       2696 CAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAACCT 2755
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138902 CAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAACCT 138843

Qy       2756 GCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTTT 2815
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138842 GCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTTT 138783

Qy       2816 ATTTACTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGAT 2875
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138782 ATTTACTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGAT 138723

Qy       2876 CTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATATGTAAAAATTA 2935
             ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||

Db     138722 CTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATACGTAAAAATTA 138663

Qy       2936 AGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTTGTTTAT 2995
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138662 AGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTTGTTTAT 138603

Qy       2996 TTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTGA 3055
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138602 TTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTGA 138543

Qy       3056 AATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAAAC 3115
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138542 AATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAAAC 138483

Qy       3116 CTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTTAAAGGAGATAAGAGGCGTAAA 3175
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138482 CTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTTAAAGGAGATAAGAGGCGTAAA 138423

Qy       3176 GTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTCATATCTTAAAGATAAAAGGTA 3235
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138422 GTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTCATATCTTAAAGATAAAAGGTA 138363

Qy       3236 CTATTATATAACCTATACACAAGATACAGGAGAAAATATGCTTGATTTTTATTTGGCAGG 3295
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138362 CTATTATATAACCTATACACAAGATACAGGAGAAAATATGCTTGATTTTTATTTGGCAGG 138303

Qy       3296 GGGGCTAGGTTGTATGGGAGTAAAAAAAACATTGAAAATTTTTAAATTGTCCAAAGAAAC 3355
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138302 GGGGCTAGGTTGTATGGGAGTAAAAAAAACATTGAAAATTTTTAAATTGTCCAAAGAAAC 138243

Qy       3356 ATTTTAAGACTCTTTAACAAAAAAGGCCATGAGTAAATCTCTATATTAACATCACTATTT 3415
             ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     138242 ATTTTAAGACTCTTTAACAAAAAAGGCCATGAGTAAATCTCTATATTAACATTACTATTT 138183

Qy       3416 ATTTTGTTTTGGAACTGGGACATGATTCTATTTGTTATAAAATAAAATTGATGT 3469
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138182 ATTTTGTTTTGGAACTGGGACATGATTCTATTTGTTATAAAATAAAATTGATGT 138129


RESULT 10
AC127071
LOCUS       AC127071               198104 bp    DNA     linear   HTG 01-AUG-2002
DEFINITION  Homo sapiens clone RP11-200K12, *** SEQUENCING IN PROGRESS ***, 2
            unordered pieces.
ACCESSION   AC127071
VERSION     AC127071.2  GI:21908383
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 198104)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
```

Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 198104)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 198104)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jul 19, 2002 this sequence version replaced gi:21743708.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: HEGN
Center clone name: RP11-200K12
---------- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 197540 bases at least Q40
Consensus quality: 197776 bases at least Q30
Consensus quality: 197873 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 3168: contig of 3168 bp in length
* 3169 3268: gap of unknown length
* 3269 198104: contig of 194836 bp in length.

FEATURES Location/Qualifiers
source 1. .198104
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-200K12"

BASE COUNT 54304 a 41678 c 42585 g 56099 t 3438 others
ORIGIN

Query Match 23.9%; Score 830.8; DB 2; Length 198104;
Best Local Similarity 99.8%; Pred. No. 2.4e-183;

Matches 832; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       2636 GGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTGTTGGC 2695
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     147690 GGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTGTTGGC 147749

Qy       2696 CAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAACCT 2755
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     147750 CAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAACCT 147809

Qy       2756 GCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTTT 2815
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     147810 GCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTTT 147869

Qy       2816 ATTTACTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGAT 2875
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     147870 ATTTACTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGAT 147929

Qy       2876 CTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATATGTAAAAATTA 2935
             ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     147930 CTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATACGTAAAAATTA 147989

Qy       2936 AGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTGTTTAT 2995
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     147990 AGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTGTTTAT 148049

Qy       2996 TTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTGA 3055
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     148050 TTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTGA 148109

Qy       3056 AATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAAAC 3115
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     148110 AATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAAAC 148169

Qy       3116 CTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTTAAAGGAGATAAGAGGCGTAAA 3175
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     148170 CTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTTAAAGGAGATAAGAGGCGTAAA 148229

Qy       3176 GTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTCATATCTTAAAGATAAAAGGTA 3235
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     148230 GTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTCATATCTTAAAGATAAAAGGTA 148289

Qy       3236 CTATTATATAACCTATACACAAGATACAGGAGAAAATATGCTTGATTTTTATTTGGCAGG 3295
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     148290 CTATTATATAACCTATACACAAGATACAGGAGAAAATATGCTTGATTTTTATTTGGCAGG 148349

Qy       3296 GGGGCTAGGTTGTATGGGAGTAAAAAAAACATTGAAAATTTTTAAATTGTCCAAAGAAAC 3355
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     148350 GGGGCTAGGTTGTATGGGAGTAAAAAAAACATTGAAAATTTTTAAATTGTCCAAAGAAAC 148409

Qy       3356 ATTTTAAGACTCTTTAACAAAAAAGGCCATGAGTAAATCTCTATATTAACATCACTATTT 3415
             ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db     148410 ATTTTAAGACTCTTTAACAAAAAAGGCCATGAGTAAATCTCTATATTAACATTACTATTT 148469

Qy       3416 ATTTTGTTTTGGAACTGGGACATGATTCTATTTGTTATAAAATAAAATTGATGT 3469
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     148470 ATTTTGTTTTGGAACTGGGACATGATTCTATTTGTTATAAAATAAAATTGATGT 148523
```

RESULT 11
BC031060
LOCUS BC031060 3006 bp mRNA linear PRI 13-JUN-2002
DEFINITION Homo sapiens, Similar to piwi like homolog 1 (Drosophila), clone MGC:33296 IMAGE:5271278, mRNA, complete cds.
ACCESSION BC031060
VERSION BC031060.1 GI:21410560
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3006)
AUTHORS Strausberg,R.

```
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 48 Row: b Column: 15
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES              Location/Qualifiers
     source           1. .3006
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="MGC:33296 IMAGE:5271278"
                      /tissue_type="Testis"
                      /clone_lib="NIH_MGC_97"
                      /lab_host="DH10B"
                      /note="Vector: pBluescript"
     CDS              4. .2562
                      /codon_start=1
                      /product="Similar to piwi like homolog 1 (Drosophila)"
                      /protein_id="AAH31060.1"
                      /db_xref="GI:21410561"
                      /translation="MSGRARVKARGIARSPSATEVGRIQASPLPRSVDLSNNEASSSN
                      GFLGTSRISTNDKYGISSGDAGSTFMERGVKNKQDFMDLSICTREKLAHVRNCKTGSS
                      GIPVKLVTNLFNLDFPQDWQLYQYHVTYIPDLASRRLRIALLYSHSELSNKAKAFDGA
                      ILFLSQKLEEKVTELSSETQRGETIKMTITLKRELPSSSPVCIQVFNIIFRKILKKLS
                      MYQIGRNFYNPSEPMEIPQHKLSLWPGFAISVSYFERKLLFSADVSYKVLRNETVLEF
                      MTALCQRTGLSCFTQTCEKQLIGLIVLTRYNNRTYSIDDIDWSVKPTHTFQKRDGTEI
                      TYVDYYKQQYDITVSDLNQPMLVSLLKKKRNDNSEAQLAHLIPELCFLTGLTDQATSD
                      FQLMKAVAEKTRLSPSGRQQRLARLVDNIQRNTNARFELETWGLHFGSQISLTGRIVP
                      SEKILMQDHICQPVSAADWSKDIRTCKILNAQSLNTWLILCSDRTEYVAESFLNCLRR
                      VAGSMGFNVDYPKIIKVQENPAAFVRAIQQYVDPDVQLVMCILPSNQKTYYDSIKKYL
                      SSDCPVPSQCVLARTLNKQGMMMSIATKIAMQMTCKLGGELWAVEIPLKSLMVVGIDV
                      CKDALSKDVMVVGCVASVNPRITRWFSRCILQRTMTDVADCLKVFMTGALNKWYKYNH
                      DLPARIIVYRAGVGDGQLKTLIEYEVPQLLSSVAESSSNTSSRLSVIVVRKKCMPRFF
                      TEMNRTVQNPPLGTVVDSEATRNEWYDFYLISQVACRGTVSPTYYNVIYDDNGLKPDH
                      MQRLTFKLCHLYYNWPGIVSVPAPCQYAHKLTFLVAQSIHKEPSLELANHLFYL"
BASE COUNT      904 a    627 c    674 g    801 t
ORIGIN

  Query Match             19.6%;  Score 679.2;  DB 9;  Length 3006;
  Best Local Similarity   56.6%;  Pred. No. 4.7e-148;
  Matches 1320;  Conservative    0;  Mismatches 1005;  Indels    9;  Gaps    3;

Qy        419 GATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGTTAAAGAATCAAAA 478
              || |||   ||| |  || |   ||| ||  | ||  | |  ||||| | | | |  |||
Db        238 GACTTTATGGATTTGAGTATCTGTACCAGAGAAAAATTGGCACATGTGAGAAATTGTAAA 297

Qy        479 ACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGACATCCCGT 538
              ||||||||    || ||    || |   |    || |||| ||      |   | |
Db        298 ACAGGTTCCAGTGGAATACCTGTGAAACTGGTTACAAACCTCTTTAACTTAGATTTTCCC 357

Qy        539 CCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGGAAGCCAGAAGA 598
              |   | |||    |||| ||||| ||  |    ||||  |||    | | | | |||||
Db        358 CAAGACTGGCAGCTATACCAGTACCATGTGACATATATTCCAGATTTAGCATCTAGAAGG 417

Qy        599 CTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCATGCTTTTGAT 658
              ||  |   ||| | |||| |   ||    || ||          ||     | || || ||
Db        418 CTGAGAATTGCTTTACTTTATAGTCATAGTGAACTTTCCAACAAAGCAAAAGCATTCGAC 477

Qy        659 GGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTTTTTAGTAAG 718
              ||  | ||  | ||| |  | ||| ||| |  ||||||| || ||  | |  ||| |
Db        478 GGTGCCATCCTTTTTCTGTCACAAAAGCTAGAAGAAAAGGTCACAGAGTTGTCAAGTGAA 537

Qy        719 ACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATGAACTTCCACCTACA 778
              || |  |  || |||  | | | ||| || |||||  | |  |  || || ||| | |
Db        538 ACTCAAAGAGGTGAGACTATAAAGATGACTATCACCCTGAAGAGGGAGCTGCCATCAAGT 597

Qy        779 TCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAAATCATGAAT 838
              || ||    ||  | ||| |||  ||||| || ||||| | | |  | ||||   ||
Db        598 TCTCCCGTGTGCATCCAGGTCTTCAATATCATCTTCAGAAAGATCCTCAAAAAGTTGTCC 657

Qy        839 TTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCAATTGATATTCCAAGTCAC 898
              || |||||||||||| || |  ||||| ||    || ||||| || |||||     |||
Db        658 ATGTACCAAATTGGACGGAACTTCTATAATCCTTCAGAGCCAATGGAAATTCCCCAGCAC 717

Qy        899 AGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAACAGCATCATG 958
              | ||     |||||||||| ||  | | |||  |      | | |||||  |  || ||
Db        718 AAATTATCCCTTTGGCCTGGGTTTGCCATTTCTGTGTCATATTTTGAAAGGAAGCTCCTG 777

Qy        959 CTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGATTTCATGTTC 1018
              |  |  |||| || ||  | |||||||| || | |||||| ||| |||| ||||||
Db        778 TTTAGTGCTGATGTGAGTTACAAAGTCCTCCGGAATGAGACGGTTCTGGAATTCATGACT 837

Qy        1019 AACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAAGAANTAATA 1078
               | | |||    || |          ||      |     |  ||  | |||||
Db        838 GCTCTCTGTCAAAGAACTGGCTTGTCCTGTTTCACCCAGACGTGTGAGAAGCAGCTAATA 897

Qy        1079 GGTTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGAGTGGATGATATTGACTGG 1138
               ||  |  |||| ||||| |  || || |  |  || |||     | ||||| |||||||||
Db        898 GGGCTCATTGTCCTTACAAGATACAATAACAGAACCTACTCCATTGATGACATTGACTGG 957

Qy        1139 GACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGGGGTCAGCTTCTTAGAA 1198
               |||  ||||    ||||||| ||||    || ||| | | | ||| ||   | ||
Db        958 TCAGTGAAGCCCACACACACCTTTCAGAAGCGGGATGGCACCGAGATCACCTATGTGGAT 1017

Qy        1199 TACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGC 1258
               |||||||  | |||| ||  |          |  | ||| |||| |||||  |  | || ||
Db        1018 TACTACAAGCAGCAGTATGATATTACTGTATCGGACCTGAATCAGCCCATGCTTGTTAGT 1077

Qy        1259 CAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATGCTCATTCCT 1318
               | |  |||  | | | |  ||    |  | |  ||  |||   || || |||
Db        1078 CTGTTAAAG---AAGAAGAGAAATGACAACAGTGAGGCTCAGCTCGCCCACCTGATACCT 1134

Qy        1319 GAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAACGTGATGAAA 1378
               |||||||||| ||| ||||| || |||||  |          ||||||  | |||||||
Db        1135 GAGCTCTGCTTTCTAACAGGGCTGACTGACCAGGCAACATCTGATTTCCAGCTGATGAAG 1194

Qy        1379 GACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTGGGACGACTC 1438
               |   | || |   | ||| | || | |||          ||||||     |||  | ||
Db        1195 GCTGTGGCTGAAAAGACACGTCTCAGTCCTTCAGGCCGGCAGCAGCGCCTGGCCAGGCTT 1254

Qy        1439 ATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGGGTTTGAGC 1498
               | ||  |||| || |  ||    |||| ||      || ||      ||||||  ||
Db        1255 GTGGACAACATCCAGAGGAATACCAATGCTCGCTTTGAACTAGAGACCTGGGGACTGCAT 1314

Qy        1499 TTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAGATTCACCAA 1558
               ||||   | |  ||  ||  |  | ||  | ||| |||   |||||||  ||
Db        1315 TTTGGAAGCCAGATA---TCTCTGACTGGCCGGATTGTGCCTTCAGAAAAAATATTAATG 1371

Qy        1559 GGTGGAAAAACATTTGATTACAATCCCACAATTTGCAGATTGGTCCAAAGAAACAAGAGGT 1618
               |   | | ||   | |  |          ||| || |||||||| || |   ||  |
Db        1372 CAAGACCACATAT---GTCAACCTGTGTCTGCTGCTGACTGGTCCAAGGATATTCGAACT 1428

Qy        1619 GCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACGCGAAGAAAT 1678
               |  | | ||  || |  |  ||| |||| || |  | | ||     |||| |
Db        1429 TGCAAGATTTTAAATGCACAGTCTTTGAATACCTGGTTGATTTTATGTAGCGACAGAACT 1488

Qy        1679 TATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGCATG 1738
```

```
Db       1489 GAATATGTTGCCGAGAGCTTTCTGAACTGCTTGAGAAGAGTTGCAGGTTCCATGGGATTT 1548

Qy       1739 CAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACTGAAGCCTACTTAAGAGTC 1798

Db       1549 AATGTGGACTACCCCAAAATCATAAAAGTACAAGAAAATCCAGCTGCATTTGTTAGAGCT 1608

Qy       1799 TTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCTGTTGTCAAGTAATCGG 1858

Db       1609 ATACAGCAATATGTTGATCCTGATGTTCAGCTGGTAATGTGCATTCTGCCTTCTAATCAG 1668

Qy       1859 AAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAG 1918

Db       1669 AAGACCTATTATGATTCCATTAAAAAATATTTGAGCTCAGACTGCCCAGTCCCAAGCCAA 1728

Qy       1919 TGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAAAGATT 1978

Db       1729 TGTGTGCTTGCTCGGACCTTGAATAAACAGGGCATGATGATGAGTATCGCCACCAAGATC 1788

Qy       1979 GCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAG 2038

Db       1789 GCTATGCAGATGACTTGCAAGCTCGGAGGCGAGCTGTGGGCTGTGGAAATACCTTTAAAG 1848

Qy       2039 CTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATC 2098

Db       1849 TCCCTGATGGTGGTCGGTATTGATGTCTGTAAAGATGCACTCAGCAAGGACGTGATGGTT 1908

Qy       2099 GCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCATATTT 2158

Db       1909 GTTGGATGCGTGGCAAGTGTTAACCCCAGAATCACCAGGTGGTTTTCCCGCTGTATCCTT 1968

Qy       2159 CAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGGCTCTGAGG 2218

Db       1969 CAGAGAACAATGACTGATGTTGCAGATTGCTTGAAAGTTTTCATGACTGGAGCACTCAAC 2028

Qy       2219 GCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATCGTGTACCGCGATGGCGTA 2278

Db       2029 AAATGGTACAAGTACAATCATGATTTGCCAGCACGGATAATTGTGTACCGTGCTGGTGTA 2088

Qy       2279 GGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCACAGTTTTTGGATTGTCTA 2338

Db       2089 GGGGATGGTCAGCTGAAAACACTTATTGAATATGAAGTCCCACAGCTGCTGAGCAGTGTG 2148

Qy       2339 AAATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATTGTGGTGAAGAAAAGAGTG 2398

Db       2149 GCAGAATCCAGCTCAAATACCAGCTCAAGACTGTCGGTGATTGTGGTCAGGAAGAAGTGC 2208

Qy       2399 AACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTCCTGGAACAGTT 2458

Db       2209 ATGCCACGATTCTTTACCGAAATGAACCGCACTGTACAGAACCCCCCACTTGGCACTGTT 2268

Qy       2459 ATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTTATCGTGAGCCAGGCTGTG 2518

Db       2269 GTGGATTCAGAAGCAACACGTAACGAATGGTATGACTTTTATCTGATCAGCCAGGTGGCC 2328

Qy       2519 AGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTATGACAACAGCGGCCTGAAG 2578

Db       2329 TGCCGGGGAACTGTTAGTCCTACCTACTATAATGTCATCTATGATGACAACGGCTTGAAG 2388

Qy       2579 CCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATCTATTACAACTGGCCAGGT 2638

Db       2389 CCCGACCATATGCAGAGACTTACATTCAAATTGTGCCACCTGTACTACAACTGGCCGGGC 2448

Qy       2639 GTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTGTTGGCCAG 2698

Db       2449 ATAGTCAGTGTCCCAGCACCATGTCAGTATGCTCACAAGCTGACCTTTCTGGTGGCACAA 2508

Qy       2699 AGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAA 2752

Db       2509 AGCATTCATAAAGAACCCAGTCTGGAATTAGCCAACCATCTCTTCTACCTGTGA 2562
```

RESULT 12

```
AK093475
LOCUS       AK093475               3138 bp    mRNA    linear   PRI 15-JUL-2002
DEFINITION  Homo sapiens cDNA FLJ36156 fis, clone TESTI2025546, weakly similar
            to Mus musculus mRNA for MIWI (piwi).
ACCESSION   AK093475
VERSION     AK093475.1  GI:21752360
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens testis cDNA to mRNA, clone_lib:TESTI2
            clone:TESTI2025546.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S.,
            Ono,Y., Hotuta,T., Hiraoka,S., Murakawa,K., Takiguchi,S.,
            Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M.,
            Yamashita,H., Chiba,Y., Sugiyama,T., Irie,R., Otsuki,T., Sato,H.,
            Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y.,
            Saito,K., Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y.,
            Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3138)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1. .3138
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="TESTI2025546"
                     /tissue_type="testis"
                     /clone_lib="TESTI2"
                     /note="cloning vector: pME18SFL3"
     CDS             152. .2710
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAC04179.1"
                     /db_xref="GI:21752361"
                     /translation="MSGRARVKARGIARSPSATEVGRIQASPLPRSVDLSNNEASSSN
                     GFLGTSRISTNDKYGISSGDAGSTFMERGVKNKQDFMDLSICTREKLAHVRNCKTGSS
                     GIPVKLVTNLFNLDFPQDWQLYQYHVTYIPDLASRRLRIALLYSHSELSNEAKAFDGA
                     ILFLSQKLEEKVTELSSETQRGETIKMTITLKRELPSSSPVCIQVFNIIFRKILKKLS
                     MYQIGRNFYNPSEPMEIPQHKLSLWPGFAISVSYFERKLLFSADVSYKVLRNETVLEF
                     MTALCQRTGLSCFTQTCEKQLIGLIVLTRYNNRTYSIDDIDWSVKPTHTFQKRDGTEI
                     TYVDYYKQQYDITVSDLNQPMLVSLLKKKRNDNSEAQLAHLIPELCFLTGLTDQATSD
                     FQLMKAVAEKTRLSPSGRQQRLARLVDNIQRNTNARFELETWGLHFGSQISLTGRIVP
                     SEKILMQDHICQPVSAADWSKDIRTCKILNAQSLNTWLILCSDRTEYVAESFLNCLRR
                     VAGSMGFNVDYPKIIKVQENPAAFVRAIQQYVDPDVQLVMCILPSNQKTYYDSIKKYL
                     SSDCPVPSQCVLARTLNKQGMMMSIATKIAMQMTCKLGGELWAVEIPLKSLMVVGIDV
                     CKDALSKDVMVVGCVASVNPRITRWFSRCILQRTMTDVADCLKVFMTGALNKWYKYNH
                     DLPARIIVYRAGVGDGQLETLIEYEVPQLLSSVAESGSNTSSRLSVIVVRKKCMPRFF
                     TEMNRTVQNPPLGTVVDSEATRNEWYDFYLISQVACRGTVSPTYYNVIYDDNGLKPDH
                     MQRLTFKLCHLYYNWPGIVSVPAPCQYAHKLTFLVAQSIHKEPSLELANHLFYL"
BASE COUNT      912 a    678 c    714 g    834 t
ORIGIN
```

Query Match 19.6%; Score 679.2; DB 9; Length 3138;
Best Local Similarity 56.6%; Pred. No. 4.8e-148;
Matches 1320; Conservative 0; Mismatches 1005; Indels 9; Gaps 3;



```
Db       2537 CCCGACCATATGCAGAGACTTACATTCAAATTGTGCCACCTGTACTACAACTGGCCGGGC 2596

Qy       2639 GTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTGTTGGCCAG 2698

Db       2597 ATAGTCAGTGTCCCAGCACCATGTCAGTATGCTCACAAGCTGACCTTTCTGGTGGCACAA 2656

Qy       2699 AGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAA 2752

Db       2657 AGCATTCATAAAGAACCCAGTCTGGAATTAGCCAACCATCTCTTCTACCTGTGA 2710


RESULT 13
AK096837
LOCUS       AK096837               2456 bp    mRNA    linear   PRI 15-JUL-2002
DEFINITION  Homo sapiens cDNA FLJ39518 fis, clone PUAEN2001188, moderately
            similar to Mus musculus mRNA for MIWI (piwi).
ACCESSION   AK096837
VERSION     AK096837.1  GI:21756416
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens pulmonary artery endothelial cells (HPAEC) cDNA to
            mRNA, clone_lib:PUAEN2 clone:PUAEN2001188.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Tanigami,A., Fujiwara,T., Shibahara,T., Goto,Y., Hirao,M.,
            Shimizu,F., Wakebe,H., Ono,T., Hishigaki,H., Watanabe,T., Ozaki,K.,
            Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
            Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
            Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
            Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
            Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
            Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2456)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1. .2456
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="PUAEN2001188"
                     /cell_type="pulmonary artery endothelial cells (HPAEC)"
                     /clone_lib="PUAEN2"
                     /note="cloning vector: pME18SFL3-primary culture,
                     pulmonary artery endothelial cells"
     CDS             28. .2028
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAC04873.1"
                     /db_xref="GI:21756417"
                     /translation="MTITLKRELPSSSPVCIQVFNIIFRKILKKLSMYQIGRNFYNPS
                     EPMEIPQHKLSLWPGFAISVSYFERKLLFSADVSYKVLRNETVLEFMTALCQRTGLSC
                     FTQTCEKQLIGLIVLTRYNNRTYSIDDIDWSVKPTHTFLKRDGTEITYVDYYKQQYDI
                     TVSDLNQPMLVSLLKKKRNDNSEPQLAHLIPELCFLTGLTDQATSDFQLMKAVAEKTR
                     LSPSGRQQRLARLVDNIQRNTNARFELETWGLHFGSQISLTGRIVPSEKILMQDHICQ
                     PVSAADWSKDIRTCKILNAQSLNTWLILCSDRTEYVAESFLNCLRRVTGSMGFNVDYP
                     KIIKVQENPAAFVRAIQQYVDPDVQLVMCILPSNQKTYYDSIKKYLSSDCPVPSQCVL
                     ARTLNKQGMMMSIATKIAMQMTCKLGGELWAVEIPLKSLMVVGIDVCKDALSKDVMVV
                     GCVASVNPRITRWFSRCILQRTMTDVADCLKVFMTGALNKWYKYNHDLPARIIVYRAG
                     VGDGQLKTLIEYEVPQLLSSVAESSSNTSSRLSVIVVRKKCMPRFFTEMNRTVQNPPL
```

```
                     GTVVDSEATRNEWYDFYLISQVACRGTVSPTYYNVIYDDNGLKPDHMQRLTFKLCHLY
                     YNWPGIVSVPAPCQYAHKLTFLVAQSIHKEPSLELANHLFYL"
BASE COUNT      718 a    519 c    550 g    669 t
ORIGIN

  Query Match           17.7%;  Score 613.2;  DB 9;  Length 2456;
  Best Local Similarity 57.3%;  Pred. No. 1.3e-132;
  Matches 1167;  Conservative    0;  Mismatches  860;  Indels    9;  Gaps    3;

Qy        717 AGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATGAACTTCCACCTA 776

Db          2 AAACTCAAAGAGGTGAGACTATAAAGATGACTATCACCCTGAAGAGGGAGCTGCCATCAA 61

Qy        777 CATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAAATCATGA 836

Db         62 GTTCTCCCGTGTGCATCCAGGTCTTCAATATCATCTTCAGAAAGATCCTCAAAAAGTTGT 121

Qy        837 ATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCAATTGATATTCCAAGTC 896

Db        122 CCATGTACCAAATTGGACGGAACTTCTATAATCCTTCAGAGCCAATGGAAATTCCCCAGC 181

Qy        897 ACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAACAGCATCA 956

Db        182 ACAAATTATCCCTTTGGCCTGGGTTCGCCATTTCTGTGTCATATTTTGAAAGGAAGCTCC 241

Qy        957 TGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGATTTCATGT 1016

Db        242 TGTTTAGTGCTGATGTGAGTTACAAAGTCCTCCGGAATGAGACGGTTCTGGAATTCATGA 301

Qy       1017 TCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAAGAANTAA 1076

Db        302 CTGCTCTCTGTCAAAGAACTGGCTTGTCCTGTTTCACCCAGACGTGTGAGAAGCAGCTAA 361

Qy       1077 TAGGTTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGAGTGGATGATATTGACT 1136

Db        362 TAGGGCTCATTGTCCTTACAAGATACAATAACAGAACCTACTCCATTGATGACATTGACT 421

Qy       1137 GGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGGGGTCAGCTTCTTAG 1196

Db        422 GGTCAGTGAAGCCCACACACACCTTTCTAAAGCGGGATGGCACCGAGATCACCTATGTGG 481

Qy       1197 AATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCA 1256

Db        482 ATTACTACAAGCAGCAGTATGATATTACTGTATCGGACCTGAATCAGCCCATGCTTGTTA 541

Qy       1257 GCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATGCTCATTC 1316

Db        542 GTCTGTTAAAG---AAGAAGAGAAATGACAACAGTGAGCCTCAGCTCGCCCACCTGATAC 598

Qy       1317 CTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAACGTGATGA 1376

Db        599 CTGAGCTCTGCTTTCTAACAGGGCTGACTGACCAGGCAACATCTGATTTCCAGCTGATGA 658

Qy       1377 AAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTGGGACGAC 1436

Db        659 AGGCTGTGGCTGAAAAGACACGTCTCAGTCCTTCAGGCCGGCAGCAGCGCCTGGCCAGGC 718

Qy       1437 TCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGGGTTTGA 1496

Db        719 TTGTGGACAACATCCAGAGGAATACCAATGCTCGCTTTGAACTAGAGACCTGGGGACTGC 778

Qy       1497 GCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAGATTCACC 1556

Db        779 ATTTTGGAAGCCAGATA---TCTCTGACTGGCCGGATTGTGCCTTCAGAAAAAATATTAA 835

Qy       1557 AAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAACAAGAG 1616

Db        836 TGCAAGACCACATAT---GTCAACCTGTGTCTGCTGCTGACTGGTCCAAGGATATTCGAA 892

Qy       1617 GTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACGCGAAGAA 1676

Db        893 CTTGCAAGATTTTAAATGCACAGTCTTTGAATACCTGGTTGATTTTATGTAGCGACAGAA 952
```

```
Qy       1677 ATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGCA 1736
Db        953 CTGAATATGTTGCCGAGAGCTTTCTGAACTGCTTGAGAAGAGTTACAGGTTCCATGGGAT 1012

Qy       1737 TGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACTGAAGCCTACTTAAGAG 1796
Db       1013 TTAATGTGGACTACCCCAAAATCATAAAAGTACAAGAAAATCCAGCTGCATTTGTTAGAG 1072

Qy       1797 TCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCTGTTGTCAAGTAATC 1856
Db       1073 CTATACAGCAATATGTTGATCCTGATGTTCAGTTGGTAATGTGCATTCTGCCTTCTAATC 1132

Qy       1857 GGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTC 1916
Db       1133 AGAAGACCTATTATGATTCCATTAAAAAATATTTGAGCTCAGACTGCCCAGTCCCAAGCC 1192

Qy       1917 AGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAAAGA 1976
Db       1193 AATGTGTGCTTGCTCGGACCTTGAATAAACAGGGCATGATGATGAGTATCGCCACCAAGA 1252

Qy       1977 TTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGA 2036
Db       1253 TCGCTATGCAGATGACTTGCAAGCTCGGAGGCGAGCTGTGGGCTGTGGAAATACCTTTAA 1312

Qy       2037 AGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAA 2096
Db       1313 AGTCCCTGATGGTGGTCGGTATTGATGTCTGTAAAGATGCACTCAGCAAGGACGTGATGG 1372

Qy       2097 TCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCATAT 2156
Db       1373 TTGTTGGATGCGTGGCCAGTGTTAACCCCAGAATCACCAGGTGGTTTTCCCGCTGTATCC 1432

Qy       2157 TTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGGCTCTGA 2216
Db       1433 TTCAGAGAACAATGACTGATGTTGCAGATTGCTTGAAAGTTTTCATGACTGGAGCACTCA 1492

Qy       2217 GGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATCGTGTACCGCGATGGCG 2276
Db       1493 ACAAATGGTACAAGTACAATCATGATTTGCCAGCACGGATAATTGTGTACCGTGCTGGTG 1552

Qy       2277 TAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCACAGTTTTTGGATTGTC 2336
Db       1553 TAGGGGATGGTCAGCTGAAAACACTTATTGAATATGAAGTCCCACAGCTGCTGAGCAGTG 1612

Qy       2337 TAAAATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATTGTGGTGAAGAAAAGAG 2396
Db       1613 TGGCAGAATCCAGCTCAAATACCAGCTCAAGACTGTCGGTGATTGTGGTCAGGAAGAAGT 1672

Qy       2397 TGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTCCTGGAACAG 2456
Db       1673 GCATGCCACGATTCTTTACCGAAATGAACCGCACTGTACAGAACCCCCCACTTGGCACTG 1732

Qy       2457 TTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTTATCGTGAGCCAGGCTG 2516
Db       1733 TTGTGGATTCAGAAGCAACACGTAACGAATGGTATGACTTTTATCTGATCAGCCAGGTGG 1792

Qy       2517 TGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTATGACAACAGCGGCCTGA 2576
Db       1793 CCTGCCGGGGAACTGTTAGTCCTACCTACTATAATGTCATCTATGATGACAACGGCTTGA 1852

Qy       2577 AGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATCTATTACAACTGGCCAG 2636
Db       1853 AGCCCGACCATATGCAGAGACTTACATTCAAATTGTGCCACCTGTACTACAACTGGCCGG 1912

Qy       2637 GTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTGTTGGCC 2696
Db       1913 GCATAGTCAGTGTCCCAGCACCATGTCAGTATGCTCACAAGCTGACCTTTCTGGTGGCAC 1972

Qy       2697 AGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAA 2752
Db       1973 AAAGCATTCATAAAGAACCCAGTCTGGAATTAGCCAACCATCTCTTCTACCTGTGA 2028
```

```
RESULT 14
AY014899
LOCUS       AY014899               3057 bp    mRNA    linear   INV 02-JAN-2001
DEFINITION  Strongylocentrotus purpuratus seawi mRNA, complete cds.
ACCESSION   AY014899
VERSION     AY014899.1  GI:12007639
KEYWORDS    .
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 3057)
  AUTHORS   Rodriguez,A.J. and Bonder,E.M.
  TITLE     Seawi - Cloning and Molecular Characterization of a Sea Urchin
            Homolog of Piwi
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3057)
  AUTHORS   Rodriguez,A.J. and Bonder,E.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-2000) Biological Sciences, Rutgers
            University-Newark, 101 Warren Street Room 135, Newark, NJ 07102,
            USA
FEATURES             Location/Qualifiers
     source          1. .3057
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
     CDS             181. .2745
                     /codon_start=1
                     /product="seawi"
                     /protein_id="AAG42533.1"
                     /db_xref="GI:12007640"
                     /translation="MDRRPGEQAPPPMQQAAPPQEASSGMPVPGRGRSRGGTTQPGPG
                     AVPKQEPRGELTSQIAKLAIGASGEGRVGGAVQDRRRRRDEIELEPTTRPENFVKQAI
                     AGDKIALIANGFKLKTKPDWQLYQYRVDFEPEILNPRARFALLKGHSALLGKGLTLDM
                     DTMYSLFKLAEKVTNLSAERKDGSNVNICVTHVATLNPMAPNTLHLYNVLFRRCLKMI
                     NMEQVGRNYYDPTAAIDIKQHGLQLWPGFVTSILQYEYDVMLLSDISHKVLRTQTVWE
                     VMNDLFNKARGRFKEEITKLMIGQIVLTKYNNKTYRVDDIDFDTTPADTFETRSGPVS
                     YVDYFKKSYERVIHDVNQPMLVSRPKKREEKKGVGPAYLPPELCFLTGLSDDMRADFN
                     VMKDMGQHTRVGPQDRCRTLSGFIKKLSSNEEVKTYLDSWGMEFDKEQVKLTGRVLPP
                     EKLFQRGKQFSYNPSNADWSRDTRGNALTDAKILNNWKIFYTRRDANRGQDFIKSLVR
                     VANPMGMNVRGPEIVELPDDRTETYTRSLQAQIAQDTQIVVVILPTNRKDRYDAIKKT
                     CVVTHPCPSQVIVSRTLSKQQMLMSVATKIAMQMNCKMGGDLWRVEIPLSNLMIIGID
                     SYHDSLTKGRSVLGFVASMNKSQTSFFSSCAFQHAQGEFGANLSTLMNNALKRYYQIN
                     EKFPERIIIFRDGVGDSQVNLVVDYELKQIKDTLDKVYPQGTVHKLAVVVVKKRINNR
                     FFANLRGGLSNPPPGTVIDDVVTKPHLYDYFIISQSVRQGSVSPTSYNVVYDTTGLKP
                     DHMQRLTYKLTHLYFNWPGTVRVPAPCMYAHKLAFLIGQSVHEVPNNRLADTLFYL"
BASE COUNT      822 a    806 c    796 g    633 t
ORIGIN

  Query Match           16.3%;  Score 565;  DB 3;  Length 3057;
  Best Local Similarity  54.8%;  Pred. No. 2.6e-121;
  Matches 1230;  Conservative    0;  Mismatches  997;  Indels   18;  Gaps    5;

Qy        513 CTAACCATTTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACT 572
Db        512 CAAACGGCTTCAAGTTGAAGACTAAACCAGACTGGCAACTCTACCAGTACCGGGTTGACT 571

Qy        573 ATAACCCACTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATC 632
Db        572 TTGAGCCTGAGATCCTGAACCCTAGAGCTCGTTTGCCCTCCTGAAGGGCCACAGTGCTC 631

Qy        633 TAATTGGAAAGTGTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGC 692
Db        632 TTCTTGGTAAAGGGTTAACGCTCGATATGGACACCATGTATTCTCTTTTCAAGCTAGCAG 691

Qy        693 AAAAGGTTACTGAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCA 752
Db        692 AAAAGGTAACCAACCTGAGCGCAGAGAGGAAGGATGGGTCTAACGTGAATATCTGTGTGA 751

Qy        753 CTTTAACAAATGAACTTCCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTT 812
Db        752 CTCATGTTGCTACTCTTAACCCCATGGCACCAAACACCCTGCATCTCTACAACGTCCTCT 811
```

```
Qy        813 TCAGGAGGCTTTTGAAAATCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAA 872
Db        812 TTAGAAGGTGTTTGAAGATGATCAATATGGAGCAAGTTGGACGCAACTACTATGATCCTA 871

Qy        873 ATGACCCAATTGATATTCCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCA 932
Db        872 CGGCCGCCATTGACATCAAGCAGCATGGGCTGCAGCTGTGGCCAGGGTTCGTGACATCCA 931

Qy        933 TCCTTCAGTATGAAAACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAA 992
Db        932 TCCTTCAGTACGAGTACGACGTGATGCTGCTGTCTGACATCAGCCACAAGGTTCTGCGTA 991

Qy        993 GTGAGACTGTTTTGGATTTCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTC 1052
Db        992 CCCAGACCGTCTGGGAGGTCATG---AATGATCTCTTCAACAAAGCTAGGGGCAGGTTCA 1048

Qy       1053 AAGAACAAGTTTCCAAAGAANTAATAGGTTTAGTTGTTCTTACCAAGTATAACATTAAGA 1112
Db       1049 AGGAGGAAATCACAAAGCTCATGATTGGACAGATTGTCCTCACAAAGTACAACAACAAGA 1108

Qy       1113 CATACAGAGTGGATGATATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCG 1172
Db       1109 CCTACCGGGTAGATGATATCGACTTTGACACTACTCCAGCGGATACTTTTGAGACCCGTT 1168

Qy       1173 ACGGCTCTGGGGTCAGCTTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCG 1232
Db       1169 CCGGACCA---GTTTCCTATGTCGATTACTTCAAGAAGTCCTATGAGAGGGTCATTCACG 1225

Qy       1233 ACTTGAAGCAGCCTGTCTTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACAC 1292
Db       1226 ACGTGAACCAACCCATGCTGGTGTCTCGCCCAAAGAAGAGGGAGGAAAAGAAGGG----- 1280

Qy       1293 TGCCAGGGCCTGCCATGCTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAA 1352
Db       1281 -AGTTGGTCCAGCCTACCTACCGCCTGAGCTGTGCTTCCTGACTGGCCTAAGCGACGACA 1339

Qy       1353 TGCGTAATGATTTTAACGTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGC 1412
Db       1340 TGCGCGCCGATTTCAACGTCATGAAGGACATGGGCCAACACACTCGTGTTGGACCCCAAG 1399

Qy       1413 AAAGGCAGCGTGAAGTGGGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAA 1472
Db       1400 ACAGATGTAGAACGCTCAGCGGCTTCATCAAAAAGCTCAGCAGCAATGAAGAAGTGAAGA 1459

Qy       1473 GGGAGCTTCGAGACTGGGGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAA 1532
Db       1460 CTTACCTAGATTCCTGGGGCATGGAGTTCGACAAAGAGCAGGTCAAACTGACTGGGCGTG 1519

Qy       1533 TTTTGCAAACAGAAAAGATTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTG 1592
Db       1520 TCCTGCCACCTGAGAAACTATTCCAGAGGGGTAAACAGTTCAGTTACAATCCGAGCAATG 1579

Qy       1593 CAGATTGGTCCAAAGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACT 1652
Db       1580 CGGACTGGTCCAGAGACACCAGGGGTAATGCCCTCACTGATGCAAAGATCCTCAACAACT 1639

Qy       1653 GGCTGTTGATCTATACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTAT 1712
Db       1640 GGAAGATCTTCTACACACGCCGGGACGCCAACAGGGGCCAGGACTTCATCAAGTCACTCG 1699

Qy       1713 TTAAAGTTACACCAGCCATGGGCATGCAAATGAGAAAAGCAATAATGATTGAAGT---GG 1769
Db       1700 TACGAGTCGCAAACCCAATGGGTATGAACGTACGAGGACCAGAGATAGTGGAACTTCCCG 1759

Qy       1770 ATGACAGAACTGAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGA 1829
Db       1760 ATGATCGTACAGAGACCTACACAAGATCCCTACAGGCACAAATCGCACAGGATACACAGA 1819

Qy       1830 TAGTTGTCTGTCTGTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACC 1889
Db       1820 TTGTGGTGGTGATCCTGCCAACGAACAGAAAGGACCGGTACGATGCCATCAAGAAGACGT 1879

Qy       1890 TGTGTACAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGC 1949
Db       1880 GCGTGGTGACCCATCCATGCCCCAGCCAGGTCATCGTCTCCAGGACCCTGTCCAAGCAGC 1939

Qy       1950 AAACTGTCATGGCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAG 2009
Db       1940 AGATGCTCATGTCTGTAGCCACCAAGATTGCCATGCAGATGAACTGCAAGATGGGCGGAG 1999

Qy       2010 AGCTCTGGAGGGTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACC 2069
Db       2000 ACCTCTGGCGAGTCGAGATCCCGCTGAGCAACCTGATGATCATCGGTATTGACAGCTACC 2059

Qy       2070 ATGACATGACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGA 2129
Db       2060 ATGACTCTCTGACCAAAGGCCGCTCCGTCCTGGGATTCGTGGCTTCCATGAACAAGTCTC 2119

Qy       2130 TGACCCGCTGGTTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGC 2189
Db       2120 AAACCAGCTTCTTCTCTAGTTGTGCATTCCAGCACGCACAGGGAGAATTCGGTGCTAACC 2179

Qy       2190 TCAAAGTCTGCCTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCA 2249
Db       2180 TGAGTACCCTCATGAACAATGCCCTGAAGCGATACTATCAGATCAACGAGAAGTTTCCTG 2239

Qy       2250 GCCGGATCATCGTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACT 2309
Db       2240 AGCGTATCATCATCTTCCGTGATGGTGTTGGTGACAGCCAAGTCAATCTGGTCGTTGACT 2299

Qy       2310 ACGAAGTGCCACAGTTTTTGGATTGTCTA---AAATCCATTGGTAGAGGTTACAACCCTA 2366
Db       2300 ATGAGCTCAAACAAATCAAGGACACCTTGGACAAAGTCTACCCTCAAGGGACGGTTCACA 2359

Qy       2367 GANTAACGGTAATTGTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAG 2426
Db       2360 AGCTGGCCGTTGTGGTCGTCAAGAAGCGCATCAATAACCGCTTCTTTGCCAATCTGAGAG 2419

Qy       2427 GAAGACTTCAGAATCCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAAT 2486
Db       2420 GAGGCCTGTCCAACCCTCCCCCTGGAACCGTCATCGATGATGTCGTCACCAAGCCCCACC 2479

Qy       2487 GGTATGACTTTTTTATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATT 2546
Db       2480 TCTATGACTACTTCATCATCAGTCAGAGTGTCCGCCAGGGCTCCGTGTCCCCCACGTCCT 2539

Qy       2547 ACAATGTCATCTATGACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACA 2606
Db       2540 ACAACGTTGTCTATGACACCACTGGTCTCAAGCCAGACCACATGCAGAGGTTGACCTACA 2599

Qy       2607 AGCTGTGCCACATCTATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGT 2666
Db       2600 AGCTGACCCATCTCTACTTCAACTGGCCGGGAACGGTGCGTGTTCCTGCCCCGTGCATGT 2659

Qy       2667 ACGCCCACAAGCTGGCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCAC 2726
Db       2660 ATGCACACAAGCTGGCATTCCTGATCGGGCAGAGCGTCCATGAAGTGCCCAACAACAGAC 2719

Qy       2727 TGTCAAACCGCCTTTACTACCTCTA 2751
Db       2720 TAGCCGACACCCTCTTCTATCTCTA 2744
```

```
RESULT 15
AX381696/c
LOCUS       AX381696                590 bp    DNA     linear   PAT 18-MAR-2002
DEFINITION  Sequence 634 from Patent WO0212280.
ACCESSION   AX381696
VERSION     AX381696.1  GI:19576518
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
```

```
  AUTHORS   Pyle,R.A., Xu,J. and Secrist,H.
  TITLE     Compositions and methods for the therapy and diagnosis of colon
            cancer
  JOURNAL   Patent: WO 0212280-A 634 14-FEB-2002;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .590
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      209 a    105 c     65 g    208 t      3 others
ORIGIN

  Query Match             15.9%;  Score 550.4;  DB 6;  Length 590;
  Best Local Similarity   99.3%;  Pred. No. 5.7e-118;
  Matches  551;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2915 TTATTTTATATGTAAAAATTAAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTT 2974
             ||||||||||||||||||||||||||||||| |||||||||||||||||||| ||| ||||
Db        590 TTATTTTATATGTAAAAATTAAGATTTTANATTTTATCTTCTTGTTTCTCANAGANATTT 531

Qy       2975 TGTGAGCATTTTTTTGTTTATTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAAC 3034
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        530 TGTGAGCATTTTTTTGTTTATTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAAC 471

Qy       3035 AGACTCTCTGAGAGTATTTGAAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGT 3094
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        470 AGACTCTCTGAGAGTATTTGAAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGT 411

Qy       3095 GAGCAAGTCCTACTTATAAACCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTT 3154
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        410 GAGCAAGTCCTACTTATAAACCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTT 351

Qy       3155 AAAGGAGATAAGAGGCGTAAAGTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTC 3214
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        350 AAAGGAGATAAGAGGCGTAAAGTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTC 291

Qy       3215 ATATCTTAAAGATAAAAGGTACTATTATATAACCTATACACAAGATACAGGAGAAAATAT 3274
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        290 ATATCTTAAAGATAAAAGGTACTATTATATAACCTATACACAAGATACAGGAGAAAATAT 231

Qy       3275 GCTTGATTTTTATTTGGCAGGGGGGCTAGGTTGTATGGGAGTAAAAAAAACATTGAAAAT 3334
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        230 GCTTGATTTTTATTTGGCAGGGGGGCTAGGTTGTATGGGAGTAAAAAAAACATTGAAAAT 171

Qy       3335 TTTTAAATTGTCCAAAGAAACATTTTAAGACTCTTTAACAAAAAAGGCCATGAGTAAATC 3394
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        170 TTTTAAATTGTCCAAAGAAACATTTTAAGACTCTTTAACAAAAAAGGCCATGAGTAAATC 111

Qy       3395 TCTATATTAACATCACTATTTATTTTGTTTTGGAACTGGGACATGATTCTATTTGTTATA 3454
             |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        110 TCTATATTAACATTACTATTTATTTTGTTTTGGAACTGGGACATGATTCTATTTGTTATA 51

Qy       3455 AAATAAAATTGATGT 3469
             |||||||||||||||
Db          50 AAATAAAATTGATGT 36


Search completed: July 22, 2003, 06:18:56
Job time : 8755.36 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2003, 11:38:12 ; Search time 692.563 Seconds
(without alignments)
11289.862 Million cell updates/sec

Title: US-09-873-737A-5
Perfect score: 3472
Sequence: 1 gggcgaattgggcccgacgt.....taaaataaaattgatgtccc 3472

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	3469	99.9	3472	21	AAA07588	Human piwi gene, d
	2	1833.6	52.8	4064	21	AAA07587	Mouse piwi gene, d
c	3	550.4	15.9	590	24	ABK55164	Human colon cancer
	4	459.4	13.2	501	24	ABK44483	cDNA encoding colo
c	5	458.4	13.2	460	22	AAS25525	Human ovarian PCR-
	6	395.4	11.4	423	21	AAF21880	Human breast and o
	7	377	10.9	3649	22	AAS13630	cDNA sequence enco
	8	329	9.5	2272	22	AAH15960	Human cDNA sequenc
	9	310.2	8.9	2838	23	ABL10571	Drosophila melanog
	10	286.4	8.2	300	20	AAZ13404	Human gene express
	11	259.8	7.5	3524	23	ABL10535	Drosophila melanog
	12	253	7.3	3047	21	AAA07586	Drosophila piwi ge
	13	238.8	6.9	1733	22	AAH14032	Human cDNA sequenc
	14	184.6	5.3	5671	23	ABL10570	Drosophila melanog
	15	170.8	4.9	548	24	ABN62610	Human cancer relat
	16	166.8	4.8	183	16	AAT23031	Human gene signatu
	17	146.6	4.2	1026	23	AAS70144	DNA encoding novel
	18	137	3.9	7667	23	ABL10534	Drosophila melanog
	19	111.8	3.2	2914	22	AAK53425	Human polynucleoti
	20	111.8	3.2	3050	22	AAH16058	Human cDNA sequenc
	21	110.6	3.2	786	22	AAH05561	Human cDNA clone (
c	22	101.6	2.9	325	22	AAI84876	Human polynucleoti
	23	95.6	2.8	444	23	AAS57327	cDNA #3 encoding p
	24	91.2	2.6	601	23	AAS70143	DNA encoding novel
	25	91	2.6	646	23	AAS77101	DNA encoding novel
	26	87.2	2.5	3996	22	AAH14510	Human cDNA sequenc
	27	83.8	2.4	2951	22	AAH24562	Translation initia
	28	82.6	2.4	642	22	AAH07789	Human cDNA clone (
	29	82.6	2.4	1968	22	AAH14966	Human cDNA sequenc
	30	79.8	2.3	516	22	AAH05278	Human cDNA clone (
	31	79.6	2.3	734	20	AAZ15059	Human gene express
	32	79.4	2.3	771	22	AAK58246	Human immune/haema
	33	78.6	2.3	213	24	ABS17735	Human genome-deriv
	34	70.2	2.0	1158	21	AAC50998	Arabidopsis thalia
	35	70.2	2.0	3195	21	AAC50992	Arabidopsis thalia
	36	70.2	2.0	3326	21	AAC47079	Arabidopsis thalia
	37	68.2	2.0	16167	24	ABL70254	Chemically treated
	38	68.2	2.0	16167	24	ABL33083	Human immune syste
	39	68.2	2.0	16167	24	ABL34529	Human metastasis a
	40	68	2.0	364	22	ABA56879	Human foetal liver
	41	68	2.0	364	22	ABA26498	Probe #4964 for ge
	42	68	2.0	364	24	ABS05149	Human genome-deriv
	43	64.6	1.9	7814	22	AAS46529	Tumour suppressor
	44	64.2	1.8	166	22	ABA69444	Human foetal liver
	45	64.2	1.8	166	22	ABA36373	Probe #14839 for g

ALIGNMENTS

RESULT 1
AAA07588
ID AAA07588 standard; DNA; 3472 BP.
XX
AC AAA07588;
XX
DT 29-AUG-2000 (first entry)
XX
DE Human piwi gene, designated hiwi.
XX
KW Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW anaemia; immunodeficiency; male infertility; human; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 167..2752
FT /*tag= a
FT /product= hiwi
FT /transl_except= (pos:392..394; aa:Xaa)
FT /transl_except= (pos:1073..1075; aa:Xaa)
FT /transl_except= (pos:2369..2370; aa:Xaa)
FT /note= "Xaa= Leu or Ile"
XX
PN WO200032039-A1.
XX
PD 08-JUN-2000.
XX
PF 03-DEC-1999; 99WO-US28764.
XX
PR 04-DEC-1998; 98US-0110901.
XX

PA (UYDU-) UNIV DUKE.
XX
PI Lin H;
XX
DR WPI; 2000-412085/35.
DR P-PSDB; AAY90235.
XX
PT Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT therapy of diseases such as cancer and in various research and
PT diagnostic applications -
XX
PS Claim 19; Page 189-194; 201pp; English.
XX
CC This sequence encodes the human piwi family protein, designated
CC hiwi. The piwi family nucleic acids and polypeptides are used in gene
CC therapy of diseases such as cancer and also in various research and
CC diagnostic applications. The sequences can also be used to treat
CC tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ Sequence 3472 BP; 1044 A; 694 C; 796 G; 935 T; 3 other;

Query Match 99.9%; Score 3469; DB 21; Length 3472;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGGCGAATTGGGCCCGACGTCGCATGCTCCCGGCCGCCATGGCGGCCGCGGGAATTCGAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGGCGAATTGGGCCCGACGTCGCATGCTCCCGGCCGCCATGGCGGCCGCGGGAATTCGAT 60

Qy         61 TCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTCTGAGGTGCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTCTGAGGTGCA 120

Qy        121 AGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAACAATGACTGGGAGAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAACAATGACTGGGAGAGC 180

Qy        181 CCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGCTGGTGGGCTCCAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGCTGGTGGGCTCCAC 240

Qy        241 TGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGCCACCAGCAGAGGGGGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGCCACCAGCAGAGGGGGA 300

Qy        301 ATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAGCCAAGTCACAAGGACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAGCCAAGTCACAAGGACT 360

Qy        361 CCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAGAGAGGAGGTCGTCGTAGAGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAGAGAGGAGGTCGTCGTAGAGA 420

Qy        421 TTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGTTAAAGAATCAAAAAC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGTTAAAGAATCAAAAAC 480

Qy        481 AGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGACATCCCGTCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGACATCCCGTCC 540

Qy        541 CCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGGAAGCCAGAAGACT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGGAAGCCAGAAGACT 600

Qy        601 CCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCATGCTTTTGATGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCATGCTTTTGATGG 660

Qy        661 AACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTTTTTAGTAAGAC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTTTTTAGTAAGAC 720

Qy        721 CCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATGAACTTCCACCTACATC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATGAACTTCCACCTACATC 780

Qy        781 ACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAAATCATGAATTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAAATCATGAATTT 840

Qy        841 GCAACAAATTGGACGAAATTATTATAACCCAAATGACCCAATTGATATTCCAAGTCACAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAACAAATTGGACGAAATTATTATAACCCAAATGACCCAATTGATATTCCAAGTCACAG 900

Qy        901 GTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAACAGCATCATGCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAACAGCATCATGCT 960

Qy        961 CTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGATTTCATGTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGATTTCATGTTCAA 1020

Qy       1021 CTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAAGAANTAATAGG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAAGAANTAATAGG 1080

Qy       1081 TTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGAGTGGATGATATTGACTGGGA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGAGTGGATGATATTGACTGGGA 1140

Qy       1141 CCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGGGGTCAGCTTCTTAGAATA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGGGGTCAGCTTCTTAGAATA 1200

Qy       1201 CTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGCCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGCCA 1260

Qy       1261 GCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATGCTCATTCCTGA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATGCTCATTCCTGA 1320

Qy       1321 GCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAACGTGATGAAAGA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAACGTGATGAAAGA 1380

Qy       1381 CTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTGGGACGACTCAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTGGGACGACTCAT 1440

Qy       1441 TGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGGGTTTGAGCTT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGGGTTTGAGCTT 1500

Qy       1501 TGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAGATTCACCAAGG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAGATTCACCAAGG 1560

Qy       1561 TGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAACAAGAGGTGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAACAAGAGGTGC 1620

Qy       1621 ACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACGCGAAGAAATTA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 ACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACGCGAAGAAATTA 1680

Qy       1681 TGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGCATGCA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGCATGCA 1740

Qy       1741 AATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACTGAAGCCTACTTAAGAGTCTT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACTGAAGCCTACTTAAGAGTCTT 1800

Qy       1801 ACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCTGTTGTCAAGTAATCGGAA 1860
```

```
Db       1801 ACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCTGTTGTCAAGTAATCGGAA 1860
Qy       1861 GGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAGTG 1920
Db       1861 GGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAGTG 1920
Qy       1921 TGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAAAGATTGC 1980
Db       1921 TGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAAAGATTGC 1980
Qy       1981 CCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCT 2040
Db       1981 CCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCT 2040
Qy       2041 CGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGC 2100
Db       2041 CGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGC 2100
Qy       2101 AGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCATATTTCA 2160
Db       2101 AGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCATATTTCA 2160
Qy       2161 GGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGGCTCTGAGGGC 2220
Db       2161 GGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGGCTCTGAGGGC 2220
Qy       2221 TTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATCGTGTACCGCGATGGCGTAGG 2280
Db       2221 TTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATCGTGTACCGCGATGGCGTAGG 2280
Qy       2281 AGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCACAGTTTTTGGATTGTCTAAA 2340
Db       2281 AGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCACAGTTTTTGGATTGTCTAAA 2340
Qy       2341 ATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATTGTGGTGAAGAAAAGAGTGAA 2400
Db       2341 ATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATTGTGGTGAAGAAAAGAGTGAA 2400
Qy       2401 CACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTCCTGGAACAGTTAT 2460
Db       2401 CACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTCCTGGAACAGTTAT 2460
Qy       2461 TGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTTATCGTGAGCCAGGCTGTGAG 2520
Db       2461 TGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTTATCGTGAGCCAGGCTGTGAG 2520
Qy       2521 AAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTATGACAACAGCGGCCTGAAGCC 2580
Db       2521 AAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTATGACAACAGCGGCCTGAAGCC 2580
Qy       2581 AGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATCTATTACAACTGGCCAGGTGT 2640
Db       2581 AGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATCTATTACAACTGGCCAGGTGT 2640
Qy       2641 CATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTGTTGGCCAGAG 2700
Db       2641 CATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTGTTGGCCAGAG 2700
Qy       2701 TATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAACCTGCAGA 2760
Db       2701 TATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAACCTGCAGA 2760
Qy       2761 AGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTTTATTTA 2820
Db       2761 AGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTTTATTTA 2820
Qy       2821 CTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGATCTAGC 2880
Db       2821 CTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGATCTAGC 2880
Qy       2881 ATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATATGTAAAAATTAAGATT 2940
Db       2881 ATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATATGTAAAAATTAAGATT 2940
Qy       2941 TTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTGTTTATTTTGA 3000
Db       2941 TTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTGTTTATTTTGA 3000
Qy       3001 AGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTGAAATGT 3060
Db       3001 AGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTGAAATGT 3060
Qy       3061 GTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAAACCTATA 3120
Db       3061 GTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAAACCTATA 3120
Qy       3121 TTAACTTTATTTTTGAGATACCTGTTTTGAATTTAAAGGAGATAAGAGGCGTAAAGTAGG 3180
Db       3121 TTAACTTTATTTTTGAGATACCTGTTTTGAATTTAAAGGAGATAAGAGGCGTAAAGTAGG 3180
Qy       3181 ATGCTCACTACAACCATAGGTGGGGTTTCAGCTCATATCTTAAAGATAAAAGGTACTATT 3240
Db       3181 ATGCTCACTACAACCATAGGTGGGGTTTCAGCTCATATCTTAAAGATAAAAGGTACTATT 3240
Qy       3241 ATATAACCTATACACAAGATACAGGAGAAAATATGCTTGATTTTTATTTGGCAGGGGGGC 3300
Db       3241 ATATAACCTATACACAAGATACAGGAGAAAATATGCTTGATTTTTATTTGGCAGGGGGGC 3300
Qy       3301 TAGGTTGTATGGGAGTAAAAAAAACATTGAAAATTTTTAAATTGTCCAAAGAAACATTTT 3360
Db       3301 TAGGTTGTATGGGAGTAAAAAAAACATTGAAAATTTTTAAATTGTCCAAAGAAACATTTT 3360
Qy       3361 AAGACTCTTTAACAAAAAAGGCCATGAGTAAATCTCTATATTAACATCACTATTTATTTT 3420
Db       3361 AAGACTCTTTAACAAAAAAGGCCATGAGTAAATCTCTATATTAACATCACTATTTATTTT 3420
Qy       3421 GTTTTGGAACTGGGACATGATTCTATTTGTTATAAAATAAAATTGATGTCCC 3472
Db       3421 GTTTTGGAACTGGGACATGATTCTATTTGTTATAAAATAAAATTGATGTCCC 3472
```

RESULT 2
AAA07587
ID AAA07587 standard; DNA; 4064 BP.
XX
AC AAA07587;
XX
DT 29-AUG-2000 (first entry)
XX
DE Mouse piwi gene, designated miwi.
XX
KW Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW anaemia; immunodeficiency; male infertility; mouse; ds.
XX
OS Mus sp.
XX

FH	Key	Location/Qualifiers
FT	CDS	191..2779
FT		/*tag= a
FT		/product= miwi
FT		/transl_except= (pos:450..452; aa:Xaa)
FT		/transl_except= (pos:1337..1339; aa:Xaa)
FT		/transl_except= (pos:2636..2638; aa:Xaa)
FT		/note= "Xaa= Leu or Ile"
FT		/transl_except= (pos:836..838; aa:Xaa)
FT		/note= "Xaa= unspecified amino acid"

XX
PN WO200032039-A1.
XX
PD 08-JUN-2000.
XX
PF 03-DEC-1999; 99WO-US28764.
XX
PR 04-DEC-1998; 98US-0110901.
XX

PA (UYDU-) UNIV DUKE.
XX
PI Lin H;
XX
DR WPI; 2000-412085/35.
DR P-PSDB; AAY90234.
XX
PT Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT therapy of diseases such as cancer and in various research and
PT diagnostic applications -
XX
PS Claim 19; Page 180-185; 201pp; English.
XX
CC This sequence encodes the mouse piwi family protein, designated
CC miwi. The piwi family nucleic acids and polypeptides are used in gene
CC therapy of diseases such as cancer and also in various research and
CC diagnostic applications. The sequences can also be used to treat
CC tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ Sequence 4064 BP; 1114 A; 978 C; 1077 G; 890 T; 5 other;

Query Match 52.8%; Score 1833.6; DB 21; Length 4064;
Best Local Similarity 81.8%; Pred. No. 0;
Matches 2125; Conservative 0; Mismatches 471; Indels 3; Gaps 1;

```
Qy       164 ACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCG 223
Db       188 AAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGAGACGGTG 247

Qy       224 CAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCG 283
Db       248 CAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACCTCAACAG 307

Qy       284 CCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAA---CAGCAGGAGGA 340
Db       308 TCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGTCGGAGCC 367

Qy       341 ACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAG 400
Db       368 ACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACTGGCAGAG 427

Qy       401 AGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGAC 460
Db       428 AGAGGAGGGCGTCGCCGAGACTTCCATGACNTTGGTGTGAACACCAGACAGAACCTTGAC 487

Qy       461 CATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCAT 520
Db       488 CATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCAC 547

Qy       521 TTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCA 580
Db       548 TTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCC 607

Qy       581 CTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGA 640
Db       608 CTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGA 667

Qy       641 AAGTGTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTT 700
Db       668 AGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGTC 727

Qy       701 ACTGAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACA 760
Db       728 ACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCACCCTGACC 787

Qy       761 AATGAACTTCCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGG 820
Db       788 AACGAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTTCAGGAGG 847

Qy       821 CTTTTGAAAATCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCA 880
Db       848 CTCTTGAAAATCATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAGTGACCCG 907

Qy       881 ATTGATATTCCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAG 940
Db       908 ATTGATATTCCAAACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCATCCTTCAG 967

Qy       941 TATGAAAACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACT 1000
Db       968 TATGAGAACAACATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAGCGAGACT 1027

Qy      1001 GTTTTGGATTTCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAA 1060
Db      1028 GTCCTAGACTTCATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCAGGAGCAA 1087

Qy      1061 GTTTCCAAAGAANTAATAGGTTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGA 1120
Db      1088 GTGTCGAAGGAGCTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGACCTACCGG 1147

Qy      1121 GTGGATGATATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCT 1180
Db      1148 GTGGATGACATTGACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGATGGCTCG 1207

Qy      1181 GGGGTCAGCTTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAG 1240
Db      1208 GAGGTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGACCTGAAG 1267

Qy      1241 CAGCCTGTCTTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGG 1300
Db      1268 CAGCCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCTGCCTGGC 1327

Qy      1301 CCTGCCATGCTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAAT 1360
Db      1328 CCAGCTATGNTCATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAATGCGCAAT 1387

Qy      1361 GATTTTAACGTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAG 1420
Db      1388 GATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCAGCGGCAG 1447

Qy      1421 CGTGAAGTGGGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTT 1480
Db      1448 CGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAGAGAGCTT 1507

Qy      1481 CGAGACTGGGGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAA 1540
Db      1508 CGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAATCTTACAA 1567

Qy      1541 ACAGAAAAGATTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGG 1600
Db      1568 TCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGCAGACTGG 1627

Qy      1601 TCCAAAGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTG 1660
Db      1628 TCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGGCTGCTG 1687

Qy      1661 ATCTATACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTT 1720
Db      1688 ATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTTCAAAGTG 1747

Qy      1721 ACACCAGCCATGGGCATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACT 1780
Db      1748 ACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGACAGAACA 1807

Qy      1781 GAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGT 1840
Db      1808 GAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGT 1867

Qy      1841 CTGTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGAT 1900
Db      1868 CTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTGTACAGAC 1927

Qy      1901 TGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATG 1960
Db      1928 TGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATG 1987

Qy      1961 GCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGG 2020
```

```
Db       1988 GCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGG 2047

Qy       2021 GTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACA 2080

Db       2048 GTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACACCACA 2107

Qy       2081 GCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGG 2140

Db       2108 GCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGG 2167

Qy       2141 TTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGC 2200

Db       2168 TTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGC 2227

Qy       2201 CTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATC 2260

Db       2228 TTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATC 2287

Qy       2261 GTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCA 2320

Db       2288 GTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCA 2347

Qy       2321 CAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATT 2380

Db       2348 CAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATC 2407

Qy       2381 GTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAAT 2440

Db       2408 GTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAAC 2467

Qy       2441 CCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTT 2500

Db       2468 CCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTC 2527

Qy       2501 ATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTAT 2560

Db       2528 ATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTAT 2587

Qy       2561 GACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATC 2620

Db       2588 GACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTG 2647

Qy       2621 TATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTG 2680

Db       2648 TACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTG 2707

Qy       2681 GCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTT 2740

Db       2708 GCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTC 2767

Qy       2741 TACTACCTCTAACCTGCAG 2759

Db       2768 TACTACCTCTAACCAGCAG 2786
```

RESULT 3
ABK55164/c
ID ABK55164 standard; cDNA; 590 BP.
XX
AC ABK55164;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human colon cancer-associated cDNA, SEQ ID No 634.
XX
KW Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200212280-A2.
XX
PD 14-FEB-2002.
XX
PF 30-JUL-2001; 2001WO-US23826.
XX
PR 03-AUG-2000; 2000US-223265P.
PR 02-OCT-2000; 2000US-237406P.
PR 20-MAR-2001; 2001US-277495P.
PR 03-JUL-2001; 2001US-302702P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Pyle RA, Xu J, Secrist H;
XX
DR WPI; 2002-257462/30.
XX
PT Novel polynucleotide encoding colon tumour polypeptides, useful as
PT vaccines for treating colon cancers -
XX
PS Claim 1; Page 329-330; 425pp; English.
XX
CC The invention relates to isolated polynucleotides (I) encoding colon
CC tumour polypeptides (II). (I) is useful for stimulating an immune
CC response in a patient and treating colon cancer in a patient.
CC Oligonucleotides derived from (I) are useful for determining the presence
CC of cancer in a patient. (I) and (II) are useful in pharmaceutical
CC compositions, e.g. vaccines, and other compositions for the diagnosis
CC and treatment of colon cancer. A composition comprising a first component
CC selected from physiologically acceptable carriers and immunostimulants,
CC and an antigen-presenting cell expressing (II) is useful for inhibiting
CC development of cancer in a patient. (I) is useful in the design and
CC preparation of ribozyme molecules for inhibiting expression of tumour
CC polypeptides and (I). ABK54531-ABK55464 represent human colon cancer cDNA
CC sequences of the invention.
XX
SQ Sequence 590 BP; 209 A; 105 C; 65 G; 208 T; 3 other;

Query Match 15.9%; Score 550.4; DB 24; Length 590;
Best Local Similarity 99.3%; Pred. No. 9.3e-128;
Matches 551; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy       2915 TTATTTTATATGTAAAAATTAAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTT 2974

Db        590 TTATTTTATATGTAAAAATTAAGATTTTANATTTTATCTTCTTGTTTCTCANAGANATTT 531

Qy       2975 TGTGAGCATTTTTTTGTTTATTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAAC 3034

Db        530 TGTGAGCATTTTTTTGTTTATTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAAC 471

Qy       3035 AGACTCTCTGAGAGTATTTGAAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGT 3094

Db        470 AGACTCTCTGAGAGTATTTGAAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGT 411

Qy       3095 GAGCAAGTCCTACTTATAAACCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTT 3154

Db        410 GAGCAAGTCCTACTTATAAACCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTT 351

Qy       3155 AAAGGAGATAAGAGGCGTAAAGTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTC 3214

Db        350 AAAGGAGATAAGAGGCGTAAAGTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTC 291

Qy       3215 ATATCTTAAAGATAAAAGGTACTATTATATAACCTATACACAAGATACAGGAGAAAATAT 3274

Db        290 ATATCTTAAAGATAAAAGGTACTATTATATAACCTATACACAAGATACAGGAGAAAATAT 231

Qy       3275 GCTTGATTTTTATTTGGCAGGGGGGCTAGGTTGTATGGGAGTAAAAAAAACATTGAAAAT 3334

Db        230 GCTTGATTTTTATTTGGCAGGGGGGCTAGGTTGTATGGGAGTAAAAAAAACATTGAAAAT 171

Qy       3335 TTTTAAATTGTCCAAAGAAACATTTTAAGACTCTTTAACAAAAAAGGCCATGAGTAAATC 3394

Db        170 TTTTAAATTGTCCAAAGAAACATTTTAAGACTCTTTAACAAAAAAGGCCATGAGTAAATC 111

Qy       3395 TCTATATTAACATCACTATTTATTTTGTTTTGGAACTGGGACATGATTCTATTTGTTATA 3454
```

```
Db        110 TCTATATTAACATTACTATTTATTTTGTTTTGGAACTGGGACATGATTCTATTTGTTATA 51

Qy       3455 AAATAAAATTGATGT 3469
              |||||||||||||||
Db         50 AAATAAAATTGATGT 36


RESULT 4
ABK44483
ID   ABK44483 standard; cDNA; 501 BP.
XX
AC   ABK44483;
XX
DT   05-JUN-2002  (first entry)
XX
DE   cDNA encoding colon tumour protein, SEQ ID No 34.
XX
KW   Human; colon tumour; vaccine; colon cancer; immunogenic;
KW   immunotherapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200212328-A2.
XX
PD   14-FEB-2002.
XX
PF   31-JUL-2001; 2001WO-US24218.
XX
PR   03-AUG-2000; 2000US-223283P.
PR   28-MAR-2001; 2001US-279763P.
PR   29-JUN-2001; 2001US-302051P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   King GE,  Meagher MJ,  Xu J,  Secrist H;
XX
DR   WPI; 2002-241739/29.
XX
PT   New colon cancer polypeptides and polynucleotides, useful as vaccines,
PT   for diagnosing, preventing, and treating colon cancer, and as markers
PT   for the progression of cancer -
XX
PS   Claim 1; SEQ ID No 34; 147pp; English.
XX
CC   The invention relates to polynucleotides encoding colon tumour proteins.
CC   The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC   compositions, such as vaccines, for the diagnosis, prevention, and
CC   treatment of colon cancer. Polynucleotide sequences may be used as
CC   hybridisation probes or primers, and in the design and preparation of
CC   ribozyme molecules for inhibiting expression of tumour polypeptides and
CC   proteins in tumour cells. The compositions are useful for stimulating an
CC   immune response against cancer, particularly for the immunotherapy of
CC   colon cancer, and as markers for the progression of cancer.
CC   ABK44450-ABK46237 represent coding sequences of human colon tumour
CC   proteins of the invention.
CC   Note: With the exception of SEQ ID No 1 and 2, the sequence data
CC   for this patent did not form part of the printed specification but was
CC   supplied by the European Patent Office.
XX
SQ   Sequence 501 BP; 134 A; 81 C; 93 G; 193 T; 0 other;

  Query Match           13.2%;  Score 459.4;  DB 24;  Length 501;
  Best Local Similarity   99.8%;  Pred. No. 5.7e-105;
  Matches  460;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2695 CCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAACC 2754
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db          1 CCAGAGTATTCACAGAGAGCCAAATCTGTCACTGGCAAACCGCCTTTACTACCTCTAACC 60

Qy       2755 TGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTT 2814
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTT 120

Qy       2815 TATTTACTTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGA 2874
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 TATTTACTTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGA 180

Qy       2875 TCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATATGTAAAAATT 2934
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 TCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATATGTAAAAATT 240

Qy       2935 AAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTGTTTA 2994
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 AAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTGTTTA 300

Qy       2995 TTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTG 3054
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 TTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTG 360

Qy       3055 AAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAAA 3114
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 AAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAAA 420

Qy       3115 CCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTTA 3155
              |||||||||||||||||||||||||||||||||||||||||
Db         421 CCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTTA 461


RESULT 5
AAS25525/c
ID   AAS25525 standard; cDNA; 460 BP.
XX
AC   AAS25525;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human ovarian PCR-subtracted cDNA library clone #1610.
XX
KW   Immunogenic protein; cancer; ovarian tumour; T-cell stimulation; ss;
KW   gene therapy; cytostatic; T-cell expansion; nucleic acid hybridisation;
KW   primer; probe.
XX
OS   Homo sapiens.
XX
PN   WO200157207-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US03733.
XX
PR   04-FEB-2000; 2000US-0180403.
PR   28-MAR-2000; 2000US-0192745.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Algate PA,  Mannion J;
XX
DR   WPI; 2001-488879/53.
XX
PT   New polynucleotides encoding ovarian tumour proteins, useful for
PT   treating ovarian cancer, and as probes, primers, and markers of cancer
PT   progression  -
XX
PS   Example 1; page 374; 378pp; English.
XX
CC   The invention comprises compositions used for the therapy and diagnosis
CC   of ovarian cancer. The compositions comprise one or more ovarian tumour
CC   proteins, their associated polynucleotides, or immunogenic portions of
CC   the proteins. The ovarian tumour polynucleotides and polypeptides are
CC   useful for stimulating and/or expanding T cells specific for a tumour
CC   protein. They are also useful for inhibiting the development of cancer in
CC   a patient with an ovarian tumour DNA or protein by incubating isolated
CC   T-cells allowing them to proliferate, and administering to the patient.
CC   The sequences can be used as markers for cancer, for example, to monitor
CC   ovarian cancer progression. Probes and primers are useful in nucleic acid
CC   hybridisation, in detecting the presence of complementary sequences in a
```

```
CC   given sample, for preparing mutant species and for preparing other
CC   genetic constructions. Sequences AAS23820-AAS25231 and AAS25328-AAS25549
CC   represent human ovarian tumour protein cDNA clones.
XX
SQ   Sequence 460 BP; 188 A; 78 C; 69 G; 125 T; 0 other;

  Query Match             13.2%;  Score 458.4;  DB 22;  Length 460;
  Best Local Similarity   99.8%;  Pred. No. 9.7e-105;
  Matches  459;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2695 CCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAACC 2754
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 CCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAACC 401

Qy       2755 TGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTT 2814
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 TGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTT 341

Qy       2815 TATTTACTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGA 2874
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        340 TATTTACTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGA 281

Qy       2875 TCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATATGTAAAAATT 2934
             |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        280 TCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATACGTAAAAATT 221

Qy       2935 AAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTGTTTA 2994
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 AAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTGTTTA 161

Qy       2995 TTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTG 3054
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        160 TTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTG 101

Qy       3055 AAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAAA 3114
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        100 AAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAAA 41

Qy       3115 CCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTT 3154
             |||||||||||||||||||||||||||||||||||||||
Db          40 CCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTT 1


RESULT 6
AAF21880
ID   AAF21880 standard; DNA; 423 BP.
XX
AC   AAF21880;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Human breast and ovarian cancer associated antigen gene SEQ ID 267.
XX
KW   Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW   nootropic; neurpprotective; antiviral; antiallergic; hepatotropic;
KW   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW   antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW   Addison's disease; allergy; autoimmune haemolytic anaemia;
KW   autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW   multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW   cardiovascular disorder; wound healing; neurological disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200055173-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05881.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-611515/58.
DR   P-PSDB; AAB58977.
XX
PT   New human breast and ovarian cancer associated gene sequences and the
PT   polypeptides encoded by these genes, useful in the prevention,
PT   treatment and diagnosis of cancer, immune disorders, cardiovascular
PT   disorders and neurological diseases -
XX
PS   Claim 1; Page 686; 1299pp; English.
XX
CC   Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC   proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC   associated with breast and ovarian cancer. Included in the invention are
CC   sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC   isolation and characterisation of the DNA and protein sequences of the
CC   invention. The breast and ovarian cancer associated DNA, protein, agonist
CC   or antagonist sequences exhibit cytostatic; immunosuppressive;
CC   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
CC   antibacterial; antifungal; antiparasitic and cardiant activity. The
CC   polynucleotide and protein sequences are used in the diagnosis of cancer,
CC   particularly breast and ovarian cancer. The nucleic acid sequences,
CC   proteins, agonists and agonists may also be used in the diagnosis,
CC   prevention and treatment of immune disorders e.g. Addison's disease,
CC   allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC   diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC   arthritis and ulcerative colitis; cardiovascular disorders such as
CC   myocardial ischaemias; wound healing; neurological diseases such as
CC   cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ   Sequence 423 BP; 104 A; 98 C; 129 G; 88 T; 4 other;

  Query Match             11.4%;  Score 395.4;  DB 21;  Length 423;
  Best Local Similarity   98.8%;  Pred. No. 5.8e-89;
  Matches  404;  Conservative    4;  Mismatches    0;  Indels    1;  Gaps    1;

Qy       1898 GATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTC 1957
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         14 GATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTC 73

Qy       1958 ATGGCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGG 2017
             ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db         74 ATGGCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGRGAGCTCTGG 133

Qy       2018 AGGGTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATG 2077
             ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||
Db        134 AGGGTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTKACCATGACATG 193

Qy       2078 ACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGC 2137
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        194 ACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGC 253

Qy       2138 TGGTTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTC 2197
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        254 TGGTTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTC 313

Qy       2198 TGCCTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATC 2257
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        314 TGCCTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATC 373

Qy       2258 ATCGTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGA 2306
             ||||||||||:| ||||||||||||||||||:|||||||||||||||
Db        374 ATCGTGTACCGSG-TGGCGTAGGAGACGGCCAGYTGAAAACACTGGTGA 421


RESULT 7
AAS13630
ID   AAS13630 standard; cDNA; 3649 BP.
XX
AC   AAS13630;
```

XX
DT 18-DEC-2001 (first entry)
XX
DE cDNA sequence encoding mammalian Spg16.
XX
KW Mammalian; reproductive-specific protein; male infertility; gene therapy;
KW spermatogenesis; sperm count disorder; anti infertility;
KW reproduction; ss.
XX
OS Mammalia.
XX
PN WO200166752-A2.
XX
PD 13-SEP-2001.
XX
PF 07-MAR-2001; 2001WO-US07371.
XX
PR 07-MAR-2000; 2000US-0187518.
PR 12-JAN-2001; 2001US-0261557.
XX
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
XX
PI Wang PJ, Page DC;
XX
DR WPI; 2001-570774/64.
DR P-PSDB; AAU07866.
XX
PT Novel reproduction-specific protein, useful for treating disorders of
PT reduced sperm count, enhancing/increasing sperm count and/or sperm
PT activity -
XX
PS Claim 1; Fig 15; 151pp; English.
XX
CC The present invention relates to the isolation of novel mammalian and
CC human reproductive-specific proteins (AAU07859-AAU07899), and the
CC nucleic acids encoding them. The nucleic acids encoding
CC reproductive-specific proteins are useful for diagnosing infertility
CC which is a result of reduced sperm count, reduced sperm motility,
CC malformed sperm or combinations of these. The sequences of the invention
CC are useful as markers for spermatogonial cells, for identifying genes or
CC proteins characteristic of male infertility, diagnosing or aiding in
CC the diagnosis of infertility in men, and for contraception in which
CC sperm production or sperm count is reduced or defective sperm is
CC produced. Antibodies to reproductive-specific proteins are useful for
CC determining the presence of these proteins in a sample obtained from a
CC man being assessed for infertility, for identifying the expression of
CC genes in particular cell type or particular developmental stage, for
CC studies of spermatogenesis, and for immunofluorescence of germ cells or
CC in Western blots for assessing the presence of the protein the antibody
CC binds. The sequences of the invention are also useful for treating
CC disorders of reduced sperm count, and for increasing sperm count and/or
CC sperm activity. The nucleic acids of the invention are useful in gene
CC therapy. AAS13623-AAS13647 represent cDNA sequences encoding for the
CC mammalian reproduction-specific proteins of the present invention.
XX
SQ Sequence 3649 BP; 947 A; 841 C; 911 G; 950 T; 0 other;

Query Match 10.9%; Score 377; DB 22; Length 3649;
Best Local Similarity 52.5%; Pred. No. 7.1e-84;
Matches 974; Conservative 0; Mismatches 862; Indels 21; Gaps 6;

```
Qy       914 CCTGGCTTCACTACTTCCATCCTTCAGTATGAAAACAGCATCATGCTCTGCACTGACGTT 973
Db         1 CCTGGCTATGCGGCTAGTATCCGGAGGACAGACGGGGGTCTCTTCCTGCTCGCTGATGTC 60

Qy       974 AGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGATTTCATGTTCAACTTTTATCATCAG 1033
Db        61 TCTCATAAGGTCATTCGGAACGACTCTGTGCTGGATGTCATGCATGCTATCTACCAGCAG 120

Qy      1034 ACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAAGAANTAATAGGTTTAGTTGTTCTT 1093
Db       121 AACAAGGAGCAC---TTCCAGGACGAGTGCAGCAAGCTTCTGGTTGGCAGCATTGTCATC 177

Qy      1094 ACCAAGTATAACATTAAGACATACAGAGTGGATGATATTGACTGGGACCAGAATCCCAAG 1153
Db       178 ACGCGCTACAACAATCGTACCTACCGAATCGATGATGTGGACTGGAACAAGACCCCTAAA 237

Qy      1154 AGCACCTTTAAGAAAGCCGACGGCTCTGGGGTCAGCTTCTTAGAATACTACAGGAAGCAA 1213
Db       238 GACAGCTTTGTCATGTCGGACGGGAAGGAAATCACATTCCTGGAATACTACAGCAAAAAC 297

Qy      1214 TACAACCAAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGCCAGCCCA---AGAGA 1270
Db       298 TATGGGATCACAGTCAAGGAAGATGACCAGCCGCTGCTGATCCACCGGCCCAGTGAGAGA 357

Qy      1271 AGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATGCTCATTCCTGAGCTCTGCTAT 1330
Db       358 CAGAATAACCATGGCATGTTGCTGAAGGGCGAGATCCTGCTGCTGCCCGAGCTCTCCTTC 417

Qy      1331 CTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAACGTGATGAAAGACTTAGCCGTT 1390
Db       418 ATGACGGGGATCCCTGAGAAGATGAAGAAGGACTTCAGGGCCATGAAGGACTTGACTCAG 477

Qy      1391 CATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTGGGACGACTCATTGATTACATT 1450
Db       478 CAGATTAACCTGAGCCCCAAGCAGCACCACGGTGCTTTGGAATGCCTGCTGCAGAGAATT 537

Qy      1451 CATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGGGTTTGAGCTTTGATTCCAAC 1510
Db       538 TCACAAAACGAGACAGCCAGCAATGAGCTGACCCGCTGGGGGCTCAGTCTGCATAAAGAT 597

Qy      1511 TTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAGATTCACCAAGGTGGAAAAACA 1570
Db       598 GTCCACAAGATTGAAGGTCGGCTTCTGCCAATGGAGAGGATCAACTTAAGGAACACTTCA 657

Qy      1571 TTTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAACAAGAGGTGCACCATTAATT 1630
Db       658 TTTGT---CACATCGGAGGGCCTGAACTGGGTTAAGGAAGTGACCAGAGATGCTTCCATT 714

Qy      1631 AGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACGCGAAGAAATTATGAAGCAGCC 1690
Db       715 CTAACTATTCCCATGCATTTCTGGGCACTCTTTTATCCAAAGAGAGCAATGGACCAAGCC 774

Qy      1691 AATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGCATGCAAATGAGAAAA 1750
Db       775 AGAGAACTGGTTAACATGTTGGAAAAGATTGCCGGGCCCATTGGCATGCGCACAAGCCCC 834

Qy      1751 GCAATAATGATTGAAGT---GGATGACAGAACTGAAGCCTACTTAAGAGTCTTACAGCAA 1807
Db       835 CCAGCCTGGGTTGAGCTGAAGGATGACCGAATAGAGACCTATATCAGGACCATTCAGTCC 894

Qy      1808 AAGGTCACAGCAGAC------ACCCAGATAGTTGTCTGTCTGTTGTCAAGTAATCGGAAG 1861
Db       895 TTACTGGGAGTTGAGGGGAAGATACAAATGGTCGTTTGCATCATCATGGGCACACGTGAT 954

Qy      1862 GACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAGTGT 1921
Db       955 GATCTCTATGGAGCCATCAAGAAGCTGTGCTGCGTGCAGTCCCCAGTGCCCTCACAGGTC 1014

Qy      1922 GTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAAAGATTGCC 1981
Db       1015 ATCAATGTCCGAACCATTGGTCAGCCCACCAGGCTTCGGAGCGTGGCTCAGAAAATTTTA 1074

Qy      1982 CTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCTC 2041
Db       1075 CTTCAGATGAACTGTAAACTGGGTGGTGAGCTCTGGGGAGTGGATATTCCGCTGAAACAA 1134

Qy      2042 GTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGCA 2101
Db       1135 CTAATGGTGATTGGAATGGATGTGTACCATGACCCCAGCAGAGGCATGCGCTCTGTGGTC 1194

Qy      2102 GGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCATATTTCAG 2161
Db       1195 GGCTTCGTGGCCAGCATAAATCTCACACTCACCAAATGGTACTCGAGGGTGGTGTTCCAG 1254

Qy      2162 GATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGGCTCTGAGGGCT 2221
```

```
Db       1255 ATGCCACATCAGGAGATTGTGGACAGCCTGAAGCTCTGCCTGGTGGGTTCCTTGAAAAAG 1314

Qy       2222 TGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATCGTGTACCGCGATGGCGTAGGA 2281

Db       1315 TATTATGAGGTGAACCATTGTCTCCCAGAGAAAATTGTGGTGTACCGAGATGGAGTGTCT 1374

Qy       2282 GACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCACAGTTTTTGGATTGTCTAAAA 2341

Db       1375 GATGGCCAGCTAAAGACAGTTGCCAACTACGAGATCCCTCAGCTGCAGAAGTGTTTTGAA 1434

Qy       2342 TCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATTGTGGTGAAGAAAAGAGTGAAC 2401

Db       1435 GCCTTTGATA---ACTACCACCCCAAGATGGTGGTGTTTGTAGTTCAGAAGAAAATCAGC 1491

Qy       2402 ACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTCCTGGAACAGTTATT 2461

Db       1492 ACCAATCTGTACCTTGCTGCTCCTGATCACTTCGTAACCCCCTCCCCCGGGACTGTGGTT 1551

Qy       2462 GATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTTATCGTGAGCCAGGCTGTGAGA 2521

Db       1552 GATCATACCATAACCAGCTGTGAGTGGGTGGATTTCTACCTTCTTGCCCATCATGTGCGA 1611

Qy       2522 AGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTATGACAACAGCGGCCTGAAGCCA 2581

Db       1612 CAGGGCTGTGGCATACCTACACACTACATCTGTGTTCTGAACACTGCAAATCTGAGCCCT 1671

Qy       2582 GACCACATACAGCGCTTGACCTACAAGCTGTGCCACATCTATTACAACTGGCCAGGTGTC 2641

Db       1672 GATCACATGCAGAGGTTGACTTTCAAACTATGCCACATGTACTGGAATTGGCCTGGTACC 1731

Qy       2642 ATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTGTTGGCCAGAGT 2701

Db       1732 ATCCGAGTTCCAGCTCCTTGCAAGTATGCCCACAAGCTAGCTTTCCTGTCCGGACAGATT 1791

Qy       2702 ATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAACCTGCA 2758

Db       1792 TTGCATCATGAGCCAGCCATCCAGCTGTGTGGGAACCTGTTCTTCCTGTAACTGGGA 1848
```

RESULT 8
AAH15960
ID AAH15960 standard; cDNA; 2272 BP.
XX
AC AAH15960;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:14558.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 14558; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2272 BP; 609 A; 509 C; 585 G; 569 T; 0 other;

Query Match 9.5%; Score 329; DB 22; Length 2272;
Best Local Similarity 51.8%; Pred. No. 6.1e-72;
Matches 871; Conservative 0; Mismatches 791; Indels 18; Gaps 5;

```
Qy       1086 TTGTTCTTACCAAGTATAACATTAAGACATACAGAGTGGATGATATTGACTGGGACCAGA 1145

Db         34 TTGTTATCACCTGATATAACAATCGTACCTATCGTATTGATGATGTGGATTGGAATAAGA 93

Qy       1146 ATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGGGGTCAGCTTCTTAGAATACTACA 1205

Db         94 CTCCAAAGGATAGCTTCACGATGTCTGATGGGAAAGAGATCACATTCTTGGAATACTACA 153

Qy       1206 GGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGCCAGCCCA 1265

Db        154 GCAAAAATTATGGGATCACAGTTAAGGAAGAGGACCAGCCATTGCTGATTCACAGGCCCA 213

Qy       1266 ---AGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATGCTCATTCCTGAGC 1322

Db        214 GTGAGAGACAGGATAATCATGGGATGCTGCTAAAAGGGGAAATCCTGCTGCTGCCTGAGC 273

Qy       1323 TCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAACGTGATGAAAGACT 1382

Db        274 TTTCTTTTATGACCGGAATCCCAGAGAAGATGAAGAAGGACTTCAGAGCCATGAAGGATT 333

Qy       1383 TAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTGGGACGACTCATTG 1442

Db        334 TGGCTCAGCAAATCAATCTGAGCCCCAAGCAACACCATAGTGCTTTGGAATGCTTGCTGC 393

Qy       1443 ATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGGGTTTGAGCTTTG 1502

Db        394 AAAGAATTGCAAAGAACGAGGCAGCCACCAATGAACTGATGCGTTGGGGGCTCCGTCTGC 453

Qy       1503 ATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAGATTCACCAAGGTG 1562

Db        454 AAAAGGATGTACATAAGATTGAAGGACGTGTTCTGCCAATGGAAAGAATTAACTTA---A 510

Qy       1563 GAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAACAAGAGGTGCAC 1622

Db        511 AAAATACTTCGTTTATCACATCTCAGGAACTAAACTGGGTTAAGGAAGTAACCAGAGACC 570

Qy       1623 CATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACGCGAAGAAATTATG 1682
```

```
Db        571 CTTCCATCTTGACTATCCCCATGCATTTCTGGGCACTTTTTTACCCAAAGAGAGCAATGG 630

Qy       1683 AAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGCATGCAAA 1742

Db        631 ACCAGGCTCGAGAACTGGTCAACATGTTGGAGAAGATAGCCGGCCCCATTGGCATGCGTA 690

Qy       1743 TGAGAAAAGCAATAATGATTGAAGT---GGATGACAGAACTGAAGCCTACTTAAGAGTCT 1799

Db        691 TGAGCCCACCGGCCTGGGTTGAACTAAAGGATGACCGAATAGAGACTTATGTCAGAACCA 750

Qy       1800 TACAGCAAAAGGTCACAGCAG------ACACCCAGATAGTTGTCTGTCTGTTGTCAAGTA 1853

Db        751 TTCAATCCACGTTAGGAGCTGAGGGGAAGATACAGATGGTTGTTTGCATCATCATGGGCC 810

Qy       1854 ATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAA 1913

Db        811 CACGTGATGATCTCTATGGGGCCATCAAGAAGCTGTGCTGTGTGCAGTCCCCAGTGCCCT 870

Qy       1914 GTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAA 1973

Db        871 CCCAGGTTGTCAATGTTCGAACCATTGGTCAGCCCACCAGGCTTCGGAGTGTGGCCCAGA 930

Qy       1974 AGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCC 2033

Db        931 AGATTTTACTTCAGATTAACTGTAAATTGGGTGGTGAGCTCTGGGGAGTGGATATTCCTC 990

Qy       2034 TGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGT 2093

Db        991 TGAAACAGTTAATGGTGATCGGGATGGATGTTTACCATGACCCCAGTAGAGGCATGCGCT 1050

Qy       2094 CAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCA 2153

Db       1051 CCGTGGTTGGCTTCGTGGCAAGCATCAATCTCACCCTCACAAAATGGTATTCCCGGGTGG 1110

Qy       2154 TATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGGCTC 2213

Db       1111 TGTTCCAGATGCCGCATCAGGAGATTGTGGACAGCCTGAAGCTATGCCTCGTGGGCTCCT 1170

Qy       2214 TGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATCGTGTACCGCGATG 2273

Db       1171 TAAAAAAGTTTTATGAGGTGAACCACTGTCTACCAGAGAAGATTGTGGTGTACCGTGATG 1230

Qy       2274 GCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCACAGTTTTTGGATT 2333

Db       1231 GAGTGTCTGATGGCCAACTGAAGACAGTTGCCAACTATGAGATTCCTCAACTACAGAAGT 1290

Qy       2334 GTCTAAAATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATTGTGGTGAAGAAAA 2393

Db       1291 GTTTTGAAGCTTTTGAGA---ATTATCAGCCCAAGATGGTGGTGTTTGTAGTTCAGAAGA 1347

Qy       2394 GAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTCCTGGAA 2453

Db       1348 AAATCAGTACTAATCTATATCTGGCTGCTCCTCAGAACTTTGTAACTCCCACTCCTGGAA 1407

Qy       2454 CAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTTATCGTGAGCCAGG 2513

Db       1408 CTGTGGTAGATCATACAATAACAAGCTGTGAGTGGGTGGATTTCTATCTTCTTGCCCATC 1467

Qy       2514 CTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTATGACAACAGCGGCC 2573

Db       1468 ATGTACGGCAGGGCTGTGGCATTCCTACGCATTATGTCTGTGTTCTCAACACCGCAAACC 1527

Qy       2574 TGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATCTATTACAACTGGC 2633

Db       1528 TGAGCCCTGATCATATGCAGAGGCTGACTTTCAAACTGTGCCACATGTACTGGAATTGGC 1587

Qy       2634 CAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTGTTG 2693

Db       1588 CTGGCACCATCAGAGTTCCAGCTCCTTGCAAGTATGCCCACAAGCTAGCTTTCCTGTCAG 1647

Qy       2694 GCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAAC 2753
```

```
Db       1648 GACACATCTTGCATCATGGACCAGCCATCCAGCTGTGCGAGAACCTGTTCTTCCTGTGAC 1707
```

RESULT 9
ABL10571
ID ABL10571 standard; cDNA; 2838 BP.
XX
AC ABL10571;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 26195.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB66468.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 26195; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2838 BP; 783 A; 659 C; 682 G; 714 T; 0 other;

Query Match 8.9%; Score 310.2; DB 23; Length 2838;
Best Local Similarity 49.1%; Pred. No. 3.6e-67;
Matches 1133; Conservative 0; Mismatches 1130; Indels 46; Gaps 10;

```
Qy        476 AAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGACATCC 535

Db        431 AAGAAAGGAGTGGTTGGCACTCATATTACCGTGCAGGCAAACTATTTTAAGGTATTAAAG 490

Qy        536 CGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGGAAGCCAGA 595

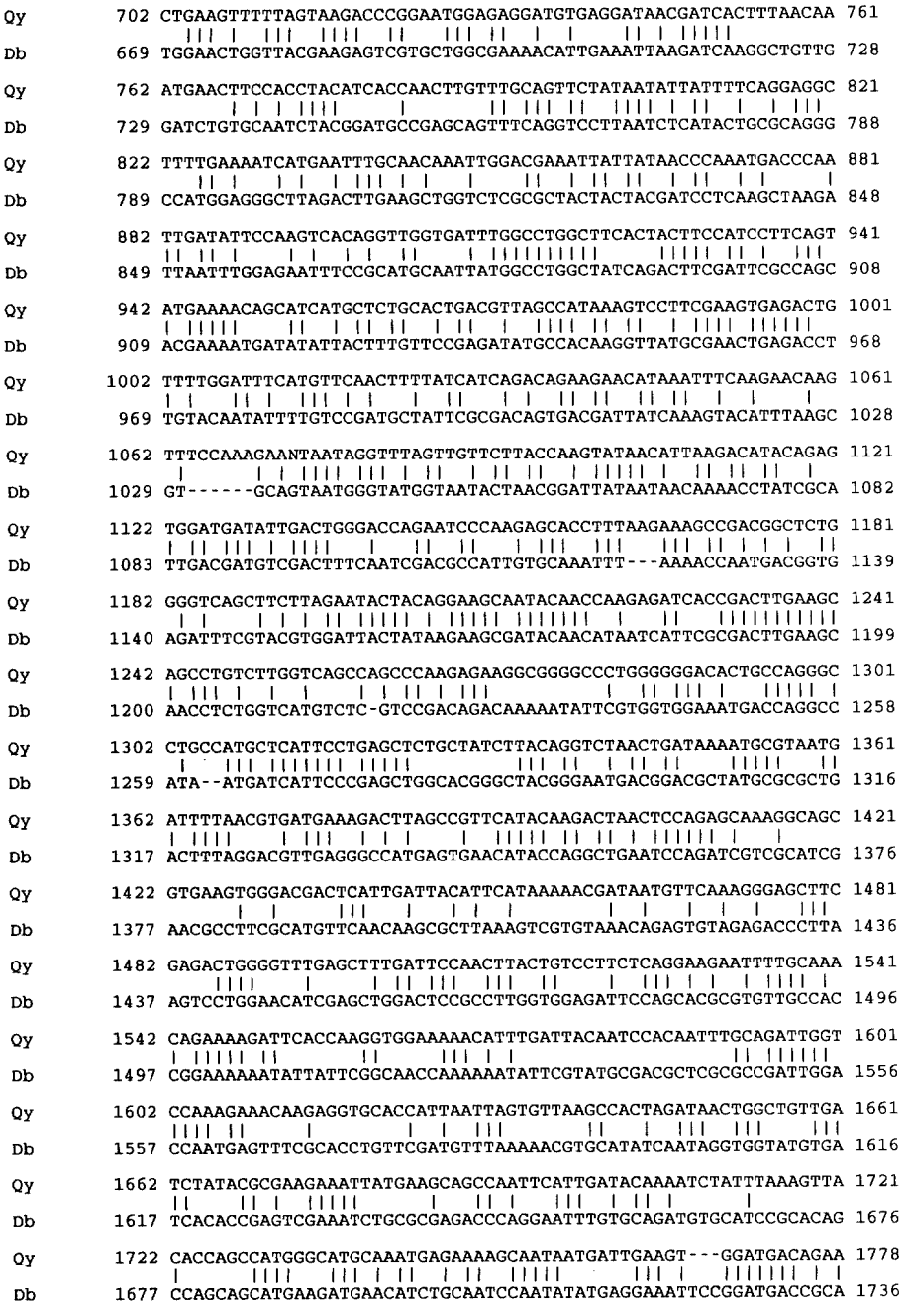
Db        491 CGTCCAAACTGGACCATCTACCAGTACCGCGTCGATTTTACGCCTGATGTGGAGGCTACA 550

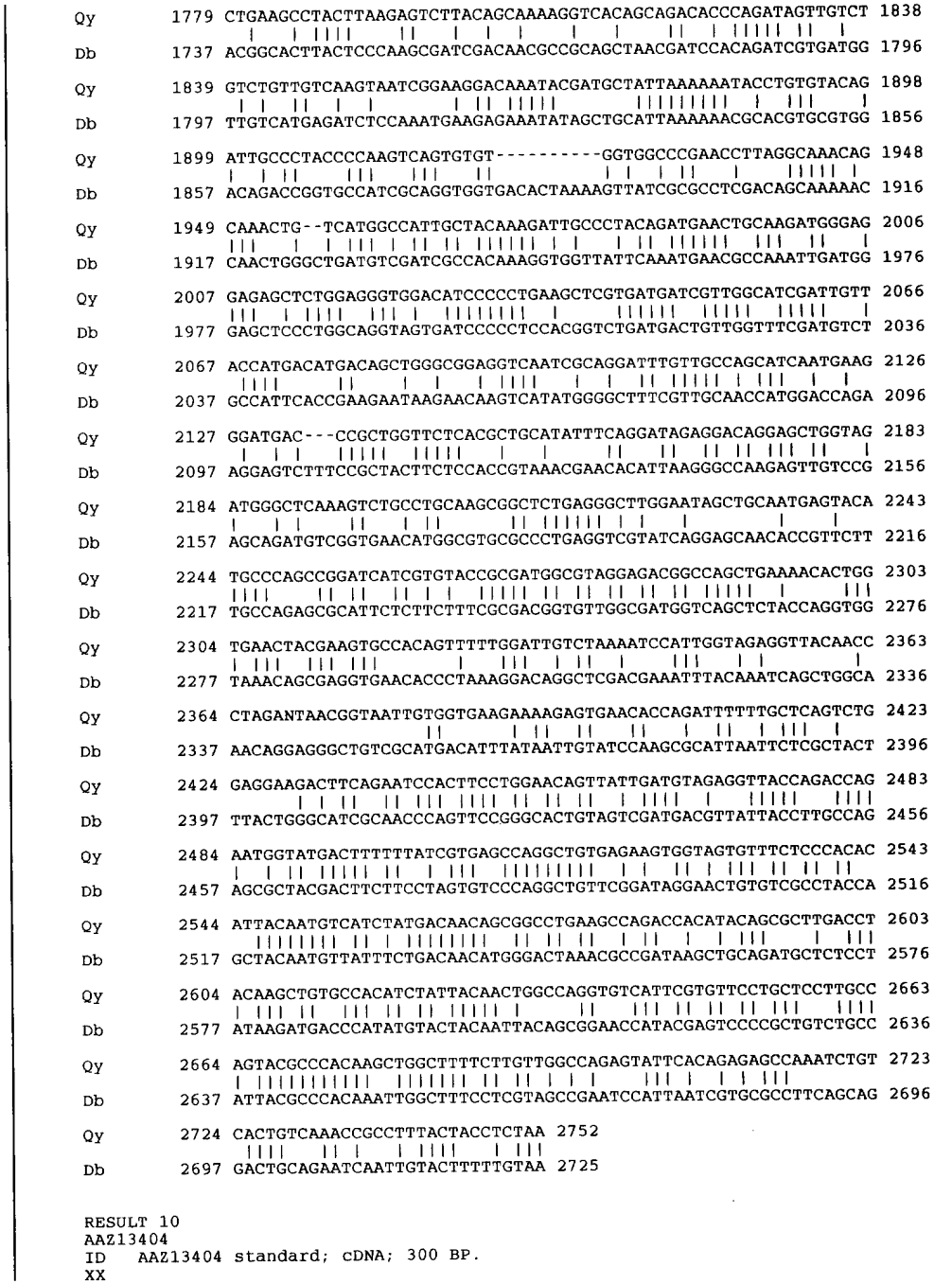
Qy        596 AGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCATGCTTTT 655

Db        551 CGACTGCGACGGTCTTTCTTGTATGAACATAAAGGGATCCTGGGC--GGCTACATCTTTG 608

Qy        656 GATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAA----------AGGTTA 701

Db        609 ACGGAACCAATATGTTTTGCATCAATCAGTTCAAAGCTGTTCAAGATAGCCCCTATGTTT 668
```





RESULT 10
AAZ13404
ID AAZ13404 standard; cDNA; 300 BP.
XX

```
AC   AAZ13404;
XX
DT   12-OCT-1999  (first entry)
XX
DE   Human gene expression product cDNA sequence SEQ ID NO:873.
XX
KW   Human; gene; gene expression product; diagnosis; therapy; probe;
KW   detection; mapping; tissue typing; profiling; forensic; cancer;
KW   genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9938972-A2.
XX
PD   05-AUG-1999.
XX
PF   28-JAN-1999;   99WO-US01619.
XX
PR   03-APR-1998;   98US-0080666.
PR   28-JAN-1998;   98US-0072910.
PR   24-FEB-1998;   98US-0075954.
PR   31-MAR-1998;   98US-0080114.
PR   03-APR-1998;   98US-0080515.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Crkvenjakov R,  Dickson M,  Drmanac R,  Drmanac S;
PI   Escobedo J,  Garcia PD,  Garcia V,  Giese K,  Innis MA;
PI   Jones WL,  Kassam A,  Kennedy GC,  Kita D,  Labat I;
PI   Lamson G,  Leshkowitz D,  Pot D,  Randazzo F,  Reinhard C;
PI   Stache-Crain B,  Sudduth-Klinger J,  Williams LT;
XX
DR   WPI; 1999-494092/41.
XX
PT   Novel human genes and their expression products which are
PT   differentially expressed in different cell types
XX
PS   Claim 1; Page 862-863; 2479pp; English.
XX
CC   The present invention describes a library of human polynucleotides
CC   comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC   a method of detecting differentially expressed genes correlated with the
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample from a cell
CC   suspected of being cancerous, where the gene product is encoded by one
CC   of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC   polynucleotides can be used as a source of primers and probes, which can
CC   be used for a variety of purpose, e.g. detection of expression levels,
CC   mapping, tissue typing or profiling, forensics, genetic analysis and
CC   detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC   can be used for raising antibodies for experimental, diagnostic and
CC   therapeutic purposes. The polynucleotides may also be used to construct
CC   arrays for diagnostics (which may be used to determine function of an
CC   encoded protein); and to detect differences in expression levels between
CC   two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC   identify a genetic predisposition or susceptibility to a disease such as
CC   cancer). The polynucleotides of the invention are especially used in the
CC   diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC   and lung cancer. The polynucleotides can also be used to screen for
CC   peptide analogues and antagonists.
XX
SQ   Sequence 300 BP; 75 A; 70 C; 91 G; 64 T; 0 other;

  Query Match             8.2%;  Score 286.4;  DB 20;  Length 300;
  Best Local Similarity   99.3%;  Pred. No. 1.1e-61;
  Matches  298;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy       1909 CCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGC 1968
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGC 60

Qy       1969 TACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACAT 2028
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACAT 120

Qy       2029 CCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCG 2088
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCG 180

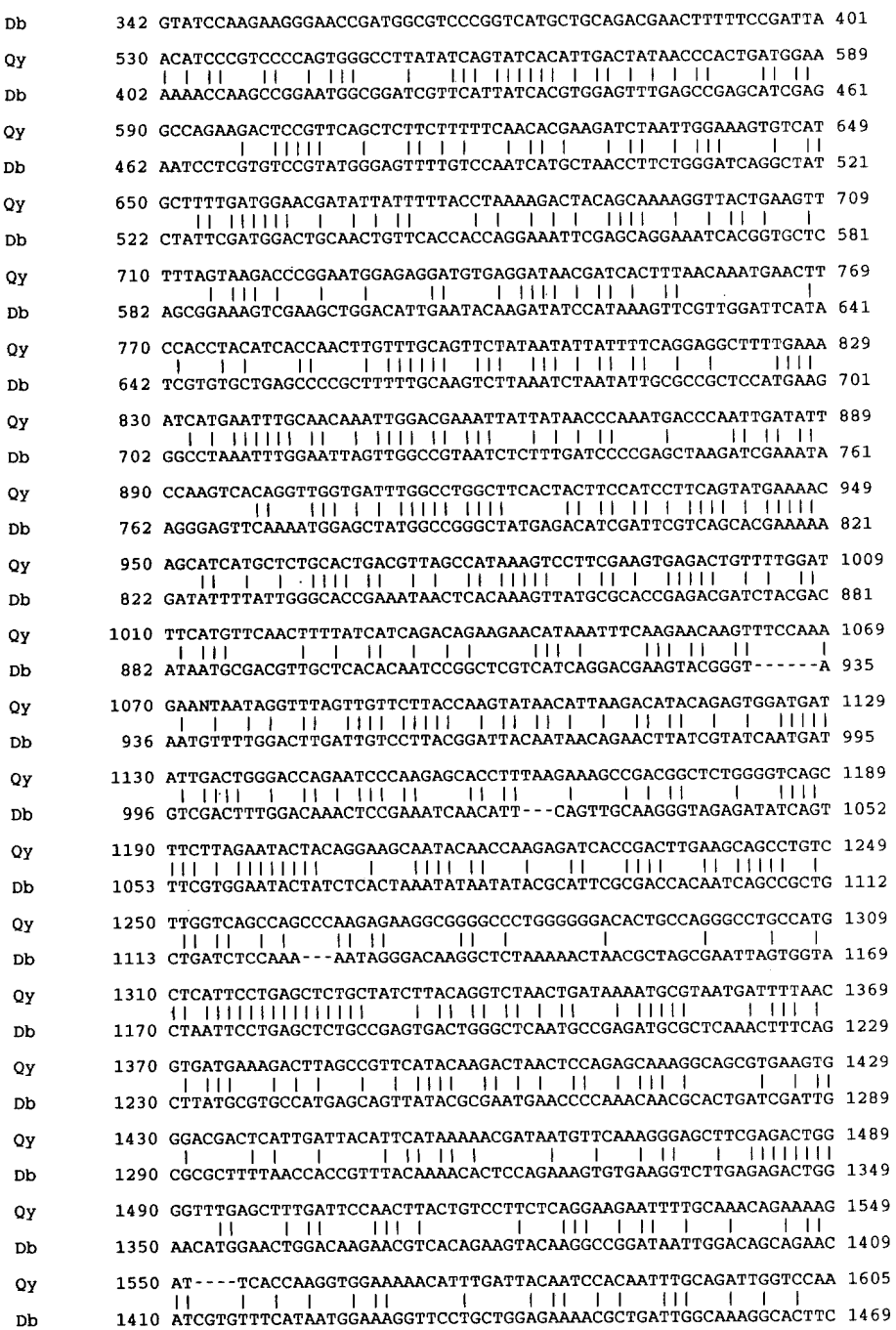
Qy       2089 GAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACG 2148
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACG 240

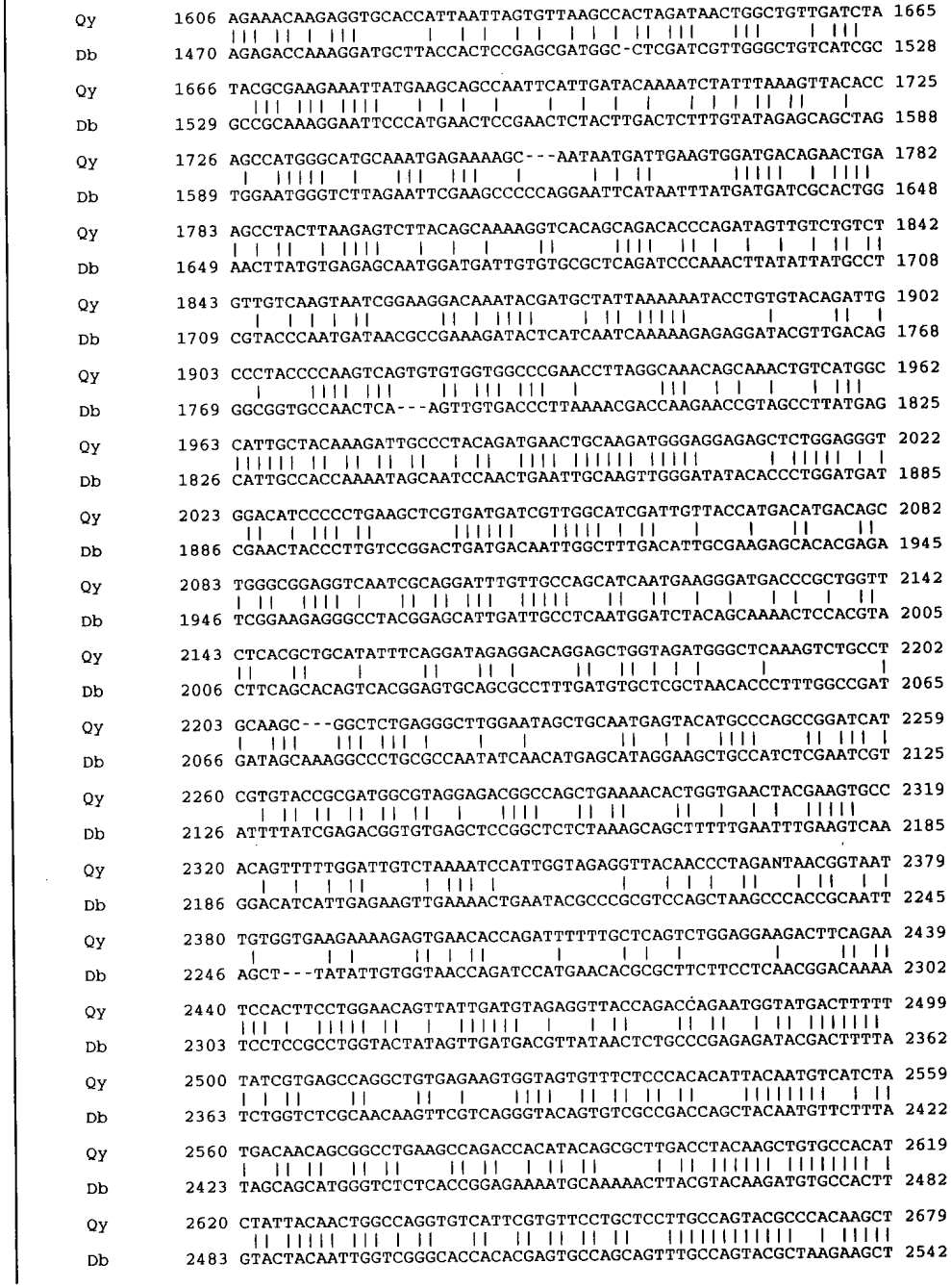
Qy       2149 CTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGC 2208
             ||||||||||||||||||||||||||||||||||||||||| || ||||||||||||
Db        241 CTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAGAG-CTGCCTGCAAGC 299


RESULT 11
ABL10535
ID   ABL10535 standard; cDNA; 3524 BP.
XX
AC   ABL10535;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 26087.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB66432.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 26087; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3524 BP; 1053 A; 741 C; 761 G; 969 T; 0 other;

  Query Match             7.5%;  Score 259.8;  DB 23;  Length 3524;
  Best Local Similarity   47.3%;  Pred. No. 1.7e-54;
  Matches 1085;  Conservative    0;  Mismatches 1179;  Indels   29;  Gaps    9;

Qy        470 GAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTG 529
              | ||| || |  ||  |    ||| |    || | | |    || |||  ||||||  |
```





```
Qy       2680 GGCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCT 2739
Db       2543 AGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTT 2602

Qy       2740 TTACTACCTCTAA 2752
Db       2603 TTATTATCTATAA 2615
```

RESULT 12
AAA07586
ID AAA07586 standard; DNA; 3047 BP.
XX
AC AAA07586;
XX
DT 29-AUG-2000 (first entry)
XX
DE Drosophila piwi gene.
XX
KW Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW anaemia; immunodeficiency; male infertility; Drosophila; ds.
XX
OS Drosophila sp.
XX
FH Key Location/Qualifiers
FT CDS 84..2615
FT /*tag= a
FT /product= piwi
FT /transl_except= (pos:120..122; aa:Xaa)
FT /transl_except= (pos:399..401; aa:Xaa)
FT /transl_except= (pos:2436..2438; aa:Xaa)
FT /note= "Xaa= Leu or Ile"
XX
PN WO200032039-A1.
XX
PD 08-JUN-2000.
XX
PF 03-DEC-1999; 99WO-US28764.
XX
PR 04-DEC-1998; 98US-0110901.
XX
PA (UYDU-) UNIV DUKE.
XX
PI Lin H;
XX
DR WPI; 2000-412085/35.
DR P-PSDB; AAY90233.
XX
PT Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT therapy of diseases such as cancer and in various research and
PT diagnostic applications -
XX
PS Claim 19; Page 171-175; 201pp; English.
XX
CC This sequence encodes the Drosophila piwi family protein, designated
CC piwi. The piwi family nucleic acids and polypeptides are used in gene
CC therapy of diseases such as cancer and also in various research and
CC diagnostic applications. The sequences can also be used to treat
CC tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ Sequence 3047 BP; 936 A; 664 C; 672 G; 772 T; 3 other;

 Query Match 7.3%; Score 253; DB 21; Length 3047;
 Best Local Similarity 47.1%; Pred. No. 8.2e-53;
 Matches 1080; Conservative 0; Mismatches 1184; Indels 29; Gaps 9;

```
Qy        470 GAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTG 529
Db        342 GTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGANTA 401

Qy        530 ACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGGAA 589
Db        402 AAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAG 461

Qy        590 GCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCAT 649
Db        462 AATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTAT 521

Qy        650 GCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTT 709
Db        522 CTATTCGACGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTC 581

Qy        710 TTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATGAACTT 769
Db        582 AGCGGAAAGTCGAAGCTGGACATTGAATACAAGATATCCATAAAGTTCGTTGGATTCATA 641

Qy        770 CCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAA 829
Db        642 TCGTGTGCTGAGCCCCGCTTTTTGCAAGTCTTAAATCTAATATTGCGCCGCTCGATGAAG 701

Qy        830 ATCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCAATTGATATT 889
Db        702 GGCCTAAATTTGGAATTAGTTGGCCGTAATCTCTTTGATCCCCGAGCTAAGATCGAAATA 761

Qy        890 CCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAAC 949
Db        762 AGGGAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTCAGCACGAAAAA 821

Qy        950 AGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGAT 1009
Db        822 GATATTTTATTGGGCACCGAAATAACTCACAAAGTTATGCGCACCGAGACGATCTACGAC 881

Qy       1010 TTCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAA 1069
Db        882 ATAATGCGACGTTGCTCACACAATCCGGCTCGTCATCAGGACGAAGTACGGGT------A 935

Qy       1070 GAANTAATAGGTTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGAGTGGATGAT 1129
Db        936 AATGTTTTGGACTTGATTGTCCTTACGGATTACAATAACAGAACTTATCGTATCAATGAT 995

Qy       1130 ATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGGGGTCAGC 1189
Db        996 GTCGACTTTGGACAAACTCCGAAATCAACATT---CAGTTGCAAGGGTAGAGATATCAGT 1052

Qy       1190 TTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTC 1249
Db        1053 TTCGTGGAATACTATCTCACTAAATATAATATACGCATTCGCGACCACAATCAGCCGCTG 1112

Qy       1250 TTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATG 1309
Db        1113 CTGATTTCCAAA---AATAGGGACAAGGCTCTAAAAACTAACGCTAGCGAATTAGTGGTA 1169

Qy       1310 CTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAAC 1369
Db        1170 CTAATTCCTGAGCTCTGCCGAGTGACTGGGCTCAATGCCGAGATGCGCTCAAACTTTCAG 1229

Qy       1370 GTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTG 1429
Db        1230 CTTATGCGTGCCATGAGCAGTTATACGCGAATGAACCCCAAACAACGCACTGATCGATTG 1289

Qy       1430 GGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGG 1489
Db        1290 CGCGCTTTTAACCACCGTTTACAAAACACTCCAGAAAGTGTGAAGGTCTTGAGAGACTGG 1349

Qy       1490 GGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAG 1549
Db        1350 AACATGGAACTGGACAAGAACGTCACAGAAGTACAAGGCCGGATAATTGGACAGCAGAAC 1409

Qy       1550 AT----TCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCCAA 1605
Db        1410 ATCGTGTTTCATAATGGAAAGGTTCCTGCTGGAGAAAACGCTGATTGGCAAAGGCACTTC 1469

Qy       1606 AGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTA 1665
Db        1470 AGAGACCAAAGGATGCTTACCACTCCGAGCGATGGC-CTCGATCGTTGGGCTGTCATCGC 1528
```

```
Qy       1666 TACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACC 1725
Db       1529 GCCGCAAAGGAATTCCCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGCTAG 1588

Qy       1726 AGCCATGGGCATGCAAATGAGAAAAGC---AATAATGATTGAAGTGGATGACAGAACTGA 1782
Db       1589 TGGAATGGGTCTTAGAATTCGAAGCCCCCAGGAATTCATAATTTATGATGATCGCACTGG 1648

Qy       1783 AGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCT 1842
Db       1649 AACTTATGTGAGAGCAATGGATGATTGTGTGCGCTCAGATCCCAAACTTATATTATGCCT 1708

Qy       1843 GTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTG 1902
Db       1709 CGTACCCAATGATAACGCCGAAAGATACTCATCAATCAAAAAGAGAGGATACGTTGACAG 1768

Qy       1903 CCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGC 1962
Db       1769 GGCGGTGCCAACTCA---AGTTGTGACCCTTAAAACGACCAAGAACCGTAGCCTTATGAG 1825

Qy       1963 CATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGT 2022
Db       1826 CATTGCCACCAAAATAGCAATCCAACTGAATTGCAAGTTGGGATATACACCCTGGATGAT 1885

Qy       2023 GGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGC 2082
Db       1886 CGAACTACCCTTGTCCGGACTGATGACAATTGGCTTTGACATTGCGAAGAGCACACGAGA 1945

Qy       2083 TGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTT 2142
Db       1946 TCGGAAGAGGGCCTACGGAGCATTGATTGCCTCAATGGATCTACAGCAAAACTCCACGTA 2005

Qy       2143 CTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCT 2202
Db       2006 CTTCAGCACAGTCACGGAGTGCAGCGCCTTTGATGTGCTCGCTAACACCCTTTGGCCGAT 2065

Qy       2203 GCAAGC---GGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCAT 2259
Db       2066 GATAGCAAAGGCCCTGCGCCAATATCAACATGAGCATAGGAAGCTGCCATCTCGAATCGT 2125

Qy       2260 CGTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCC 2319
Db       2126 ATTTTATCGAGACGGTGTGAGCTCCGGCTCTCTAAAGCAGCTTTTTGAATTTGAAGTCAA 2185

Qy       2320 ACAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAAT 2379
Db       2186 GGACATCATTGAGAAGTTGAAAACTGAATACGCCCGCGTCCAGCTAAGCCCACCGCAATT 2245

Qy       2380 TGTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAA 2439
Db       2246 AGCT---TATATTGTGGTAACCAGATCCATGAACACGCGCTTCTTCCTCAACGGACAAAA 2302

Qy       2440 TCCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTT 2499
Db       2303 TCCTCCGCCTGGTACTATAGTTGATGACGTTATAACTCTGCCCGAGAGATACGACTTTTA 2362

Qy       2500 TATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTA 2559
Db       2363 TCTGGTCTCGCAACAAGTTCGTCAGGGTACAGTGTCGCCGACCAGCTACAATGTTCTTTA 2422

Qy       2560 TGACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACAT 2619
Db       2423 TAGCAGCATGGGTNTCTCACCGGAGAAAATGCAAAAACTTACGTACAAGATGTGCCACTT 2482

Qy       2620 CTATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCT 2679
Db       2483 GTACTACAATTGGTCGGGCACCACACGAGTGCCAGCAGTTTGCCAGTACGCTAAGAAGCT 2542

Qy       2680 GGCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCT 2739
Db       2543 AGCTACCCTCGTGGGTACGAACTTGCACTCTATTCCGCAAAACGCGCTCGAAAAGAAGTT 2602

Qy       2740 TTACTACCTCTAA 2752
Db       2603 TTATTATCTATAA 2615
```

RESULT 13
AAH14032
ID AAH14032 standard; cDNA; 1733 BP.
XX
AC AAH14032;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:11143.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 11143; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 1733 BP; 435 A; 408 C; 451 G; 439 T; 0 other;

Query Match 6.9%; Score 238.8; DB 22; Length 1733;
Best Local Similarity 52.1%; Pred. No. 2.2e-49;
Matches 609; Conservative 0; Mismatches 548; Indels 12; Gaps 3;

```
Qy       1594 AGATTGGTCCAAAGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTG 1653
Db          3 AAACTGGGTTAAGGAAGTAACCAGAGACCCTTCCATCTTGACTATCCCCATGCATTTCTG 62

Qy       1654 GCTGTTGATCTATACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATT 1713
Db         63 GGCACTTTTTTACCCAAAGAGAGCAATGGACCAGGCTCGAGAACTGGTCAACATGTTGGA 122

Qy       1714 TAAAGTTACACCAGCCATGGGCATGCAAATGAGAAAAGCAATAATGATTGAAGT---GGA 1770
Db        123 GAAGATAGCCGGCCCCATTGGCATGCGTATGAGCCCACCGGCCTGGGTTGAACTAAAGGA 182

Qy       1771 TGACAGAACTGAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAG------ACAC 1824
Db        183 TGACCGAATAGAGACTTATGTCAGAACCATTCAATCCACGTTAGGAGCTGAGGGGAAGAT 242

Qy       1825 CCAGATAGTTGTCTGTCTGTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAA 1884
Db        243 ACAGATGGTTGTTTGCATCATCATGGGCCCACGTGATGATCTCTATGGGGCCATCAATAA 302

Qy       1885 ATACCTGTGTACAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAA 1944
Db        303 GCTGTGCTGTGTGCAGTCCCCAGTGCCCTCCCAGGTTGTCAATGTTCGAACCATTGGTCA 362

Qy       1945 ACAGCAAACTGTCATGGCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGG 2004
Db        363 GCCCACCAGGCTTCGGAGTGTGGCCCAGAAGATTTTACTTCAGATTAACTGTAAATTGGG 422

Qy       2005 AGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTG 2064
Db        423 TGGTGAGCTCTGGGGAGTGGATATTCCTCTGAAACAGTTAATGGTGATCGGGATGGATGT 482

Qy       2065 TTACCATGACATGACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGA 2124
Db        483 TTACCATGACCCCAGTAGAGGCATGCGCTCCGTGGTTGGCTTCGTGGCAAGCATCAATCT 542

Qy       2125 AGGGATGACCCGCTGGTTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGA 2184
Db        543 CACCCTCACAAAATGGTATTCCCGGGTGGTGTTCCAGATGCCGCATCAGGAGATTGTGGA 602

Qy       2185 TGGGCTCAAAGTCTGCCTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACAT 2244
Db        603 CAGCCTGAAGCTATGCCTCGTGGGCTCCTTAAAAAAGTTTTATGAGGTGAACCACTGTCT 662

Qy       2245 GCCCAGCCGGATCATCGTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGT 2304
Db        663 ACCAGAGAAGATTGTGGTGTACCGTGATGGAGTGTCTGATGGCCAACTGAAGACAGTTGC 722

Qy       2305 GAACTACGAAGTGCCACAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCC 2364
Db        723 CAACTATGAGATTCCTCAACTACAGAAGTGTTTTGAAGCTTTTGAGA---ATTATCAGCC 779

Qy       2365 TAGANTAACGGTAATTGTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGG 2424
Db        780 CAAGATGGTGGTGTTTGTAGTTCAGAAGAAAATCAGTACTAATCTATATCTGGCTGCTCC 839

Qy       2425 AGGAAGACTTCAGAATCCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGA 2484
Db        840 TCAGAACTTTGTAACTCCCACTCCTGGAACTGTGGTAGATCATACAATAACAAGCTGTGA 899

Qy       2485 ATGGTATGACTTTTTTATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACA 2544
Db        900 GTGGGTGGATTTCTATCTTCTTGCCCATCATGTACGGCAGGGCTGTGGCATTCCTACGCA 959

Qy       2545 TTACAATGTCATCTATGACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTA 2604
Db        960 TTATGTCTGTGTTCTCAACACCGCAAACCTGAGCCCTGATCATATGCAGAGGCTGACTTT 1019

Qy       2605 CAAGCTGTGCCACATCTATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCA 2664
Db       1020 CAAACTGTGCCACATGTACTGGAATTGGCCTGGCACCATCAGAGTTCCAGCTCCTTGCAA 1079

Qy       2665 GTACGCCCACAAGCTGGCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTC 2724
Db       1080 GTATGCCCACAAGCTAGCTTTCCTGTCAGGACACATCTTGCATCATGAACCAGCCATCCA 1139

Qy       2725 ACTGTCAAACCGCCTTTACTACCTCTAAC 2753
Db       1140 GCTGTGCGAGAACCTGTTCTTCCTGTGAC 1168
```

RESULT 14
ABL10570
ID ABL10570 standard; cDNA; 5671 BP.
XX
AC ABL10570;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 26192.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB66467.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 26192; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 5671 BP; 1723 A; 1127 C; 1131 G; 1690 T; 0 other;

Query Match 5.3%; Score 184.6; DB 23; Length 5671;
Best Local Similarity 52.8%; Pred. No. 1.6e-35;
Matches 422; Conservative 0; Mismatches 375; Indels 3; Gaps 1;

```
Qy       1956 TCATGGCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCT 2015
Db       3759 TGATGTCGATCGCCACAAAGGTGGTTATTCAAATGAACGCCAAATTGATGGGAGCTCCCT 3818

Qy       2016 GGAGGGTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACA 2075
Db       3819 GGCAGGTAGTGATCCCCCTCCACGGTCTGATGACTGTTGGTTTCGATGTCTGCCATTCAC 3878

Qy       2076 TGACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCAT---CAATGAAGGGATGA 2132
```

```
Db       3879 CGAAGAATAAGAACAAGTCATATGGGGCTTTCGTTGCAACCATGGACCAGAAGGAGTCTT 3938

Qy       2133 CCCGCTGGTTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCA 2192

Db       3939 TCCGCTACTTCTCCACCGTAAACGAACACATTAAGGGCCAAGAGTTGTCCGAGCAGATGT 3998

Qy       2193 AAGTCTGCCTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCC 2252

Db       3999 CGGTGAACATGGCGTGCGCCCTGAGGTCGTATCAGGAGCAACACCGTTCTTTGCCAGAGC 4058

Qy       2253 GGATCATCGTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACG 2312

Db       4059 GCATTCTCTTCTTTCGCGACGGTGTTGGCGATGGTCAGCTCTACCAGGTGGTAAACAGCG 4118

Qy       2313 AAGTGCCACAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGANTAA 2372

Db       4119 AGGTGAACACCCTAAAGGACAGGCTCGACGAAATTTACAAATCAGCTGGCAAACAGGAGG 4178

Qy       2373 CGGTAATTGTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGAC 2432

Db       4179 GCTGTCGCATGACATTTATAATTGTATCCAAGCGCATTAATTCTCGCTACTTTACTGGGC 4238

Qy       2433 TTCAGAATCCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATG 2492

Db       4239 ATCGCAACCCAGTTCCGGGCACTGTAGTCGATGACGTTATTACCTTGCCAGAGCGCTACG 4298

Qy       2493 ACTTTTTTATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATG 2552

Db       4299 ACTTCTTCCTAGTGTCCCAGGCTGTTCGGATAGGAACTGTGTCGCCTACCAGCTACAATG 4358

Qy       2553 TCATCTATGACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGT 2612

Db       4359 TTATTTCTGACAACATGGGACTAAACGCCGATAAGCTGCAGATGCTCTCCTATAAGATGA 4418

Qy       2613 GCCACATCTATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCC 2672

Db       4419 CCCATATGTACTACAATTACAGCGGAACCATACGAGTCCCCGCTGTCTGCCATTACGCCC 4478

Qy       2673 ACAAGCTGGCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAA 2732

Db       4479 ACAAATTGGCTTTCCTCGTAGCCGAATCCATTAATCGTGCGCCTTCAGCAGGACTGCAGA 4538

Qy       2733 ACCGCCTTTACTACCTCTAA 2752

Db       4539 ATCAATTGTACTTTTTGTAA 4558


RESULT 15
ABN62610
ID   ABN62610 standard; cDNA; 548 BP.
XX
AC   ABN62610;
XX
DT   28-JUN-2002  (first entry)
XX
DE   Human cancer related polynucleotide SEQ ID NO 2577.
XX
KW   Human; cytostatic; gene expression; gene mapping; tissue profiling;
KW   gene therapy; cancer; tumour; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200214500-A2.
XX
PD   21-FEB-2002.
XX
PF   16-AUG-2001; 2001WO-US25840.
XX
PR   16-AUG-2000; 2000US-226326P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;
PI   Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;
XX
DR   WPI; 2002-241905/29.
XX
PT   New nucleic acid for producing a polypeptide, detecting differentially
PT   expressed genes correlated with a cancerous state of a mammalian cell,
PT   and inhibiting tumor growth -
XX
PS   Claim 1; SEQ ID NO 2577; 883pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
CC   with cytostatic activity. The polynucleotide is used to produce a
CC   polypeptide, to detect differentially expressed genes correlated with a
CC   cancerous state of a mammalian cell and to inhibit tumour growth. The
CC   polynucleotide is used as a probe in mapping and tissue profiling. The
CC   encoded polypeptide and antibodies to the polypeptide can also be used
CC   for therapeutic and diagnostic purposes. The polynucleotide is useful for
CC   gene therapy.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 548 BP; 151 A; 116 C; 146 G; 134 T; 1 other;

  Query Match             4.9%;  Score 170.8;  DB 24;  Length 548;
  Best Local Similarity   57.0%;  Pred. No. 1.4e-32;
  Matches  310;  Conservative    0;  Mismatches  234;  Indels    0;  Gaps    0;

Qy       1943 AAACAGCAAACTGTCATGGCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATG 2002

Db          5 AAACAGGGCATGATGATGAGTATCGCCACCAAGATCGCTATGCAGATGACTTGCAAGCTC 64

Qy       2003 GGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGAT 2062

Db         65 GGAGGCGAGCTGTGGGCTGTGGAAATACCTTTAAAGTCCCTGATGGTGGTCGGTATTGAT 124

Qy       2063 TGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAAT 2122

Db        125 GTCTGTAAAGATGCACTCAGCAAGGACGCGATGGTTGTTGGATGCGTGGCCAGTGTTAAC 184

Qy       2123 GAAGGGATGACCCGCTGGTTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTA 2182

Db        185 CCCAGAATCACCAGGTGGTTTTCCCGCTGTATCCTTCAGAGAACAATGACTGATGTTGCA 244

Qy       2183 GATGGGCTCAAAGTCTGCCTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTAC 2242

Db        245 GATTGCTTGAAAGTTTTCATGACTGGAGCACTCAACAAATGGTACAAGTACAATCATGAT 304

Qy       2243 ATGCCCAGCCGGATCATCGTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTG 2302

Db        305 TTGCCAGCACGGATAATTGTGTACCGTGCTGGTGTAGGGGATGGTCAGCTGAAAACACTT 364

Qy       2303 GTGAACTACGAAGTGCCACAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAAC 2362

Db        365 ATTGAATATGAAGTCCCACAGCTGCTGAGCAGTGTGGCAGAATCCAGCTCANATACCAGC 424

Qy       2363 CCTAGANTAACGGTAATTGTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCT 2422

Db        425 TCAAGACTGTCGGTGATTGTGGTCAGGAAGAAGTGCATGCCACGATTCTTTACCGAAATG 484

Qy       2423 GGAGGAAGACTTCAGAATCCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCA 2482

Db        485 AACCGCACTGTACAGAACCCCCCACTTGGCACTGTTGTGGATTCAGAAGCAACACGTAAC 544

Qy       2483 GAAT 2486

Db        545 GAAT 548


Search completed: July 21, 2003, 19:01:25
Job time : 696.563 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2003, 18:06:22 ; Search time 141.728 Seconds
(without alignments)
7512.871 Million cell updates/sec

Title: US-09-873-737A-5
Perfect score: 3472
Sequence: 1 gggcgaattgggcccgacgt.....taaaataaaattgatgtccc 3472

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	60.8	1.8	2365	4	US-09-249-697A-5	Sequence 5, Appli
	2	60.8	1.8	2365	4	US-09-249-697A-18	Sequence 18, Appl
	3	60.8	1.8	2365	4	US-09-363-316B-5	Sequence 5, Appli
	4	60.8	1.8	2365	4	US-09-363-316B-23	Sequence 23, Appl
c	5	57.6	1.7	7218	1	US-08-232-463-14	Sequence 14, Appl
	6	57	1.6	1613	3	US-08-938-830-28	Sequence 28, Appl
	7	52.8	1.5	2217	4	US-09-244-314-1	Sequence 1, Appli
	8	52.8	1.5	2217	4	US-09-498-959-1	Sequence 1, Appli
	9	52.4	1.5	7218	1	US-08-232-463-14	Sequence 14, Appl
c	10	51.8	1.5	19124	2	US-08-487-826B-13	Sequence 13, Appl
c	11	51	1.5	797	1	US-08-592-126-73	Sequence 73, Appl
	12	50.4	1.5	3388	4	US-09-141-206-1	Sequence 1, Appli
	13	50.2	1.4	4203	2	US-08-866-757-1	Sequence 1, Appli
	14	50.2	1.4	4203	4	US-09-153-593-1	Sequence 1, Appli
	15	50	1.4	658	4	US-08-998-416-595	Sequence 595, App
	16	50	1.4	5852	1	US-07-867-106-2	Sequence 2, Appli
c	17	49.4	1.4	1230	4	US-09-387-574-5	Sequence 5, Appli
c	18	49.4	1.4	1230	4	US-09-668-096-5	Sequence 5, Appli
c	19	49.2	1.4	240	1	US-08-628-417-6	Sequence 6, Appli
c	20	48.8	1.4	2447	2	US-09-014-969-14	Sequence 14, Appl
	21	47	1.4	3396	4	US-08-989-299-1	Sequence 1, Appli
	22	45.6	1.3	2251	4	US-08-991-677-11	Sequence 11, Appl
c	23	45.6	1.3	6243	2	US-09-056-075-1	Sequence 1, Appli
c	24	45.4	1.3	6124	4	US-08-213-419B-3	Sequence 3, Appli
	25	44.8	1.3	163	4	US-09-182-145-51	Sequence 51, Appl
c	26	44.8	1.3	163	4	US-09-182-145-52	Sequence 52, Appl
	27	44.8	1.3	488	4	US-08-991-789A-180	Sequence 180, App
	28	44.8	1.3	488	4	US-09-062-451-180	Sequence 180, App
	29	44.8	1.3	488	4	US-09-598-326-180	Sequence 180, App
c	30	44.8	1.3	507	4	US-08-991-789A-182	Sequence 182, App
c	31	44.8	1.3	507	4	US-09-062-451-182	Sequence 182, App
c	32	44.8	1.3	507	4	US-09-598-326-182	Sequence 182, App
c	33	44.8	1.3	1222	4	US-08-861-774E-83	Sequence 83, Appl
c	34	44.8	1.3	2064	4	US-09-178-109-3	Sequence 3, Appli
c	35	44.8	1.3	2121	4	US-09-178-109-1	Sequence 1, Appli
	36	44.8	1.3	5506	4	US-09-004-838-93	Sequence 93, Appl
	37	44	1.3	44	1	US-08-381-572-5	Sequence 5, Appli
	38	44	1.3	44	1	US-08-381-572-6	Sequence 6, Appli
	39	44	1.3	44	1	US-08-592-820-5	Sequence 5, Appli
	40	44	1.3	44	1	US-08-592-820-6	Sequence 6, Appli
	41	44	1.3	44	2	US-08-670-707A-13	Sequence 13, Appl
	42	44	1.3	44	4	US-09-037-601-13	Sequence 13, Appl
	43	44	1.3	44	4	US-09-234-393-5	Sequence 5, Appli
	44	44	1.3	44	4	US-09-124-238A-19	Sequence 19, Appl
	45	44	1.3	44	4	US-09-323-873A-23	Sequence 23, Appl

ALIGNMENTS

RESULT 1
US-09-249-697A-5
; Sequence 5, Application US/09249697A
; Patent No. 6392018
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Yeung, George
; TITLE OF INVENTION: NOVEL EGF MOTIF PROTEIN OBTAINED FROM A CDNA LIBRARY OF FETAL
; TITLE OF INVENTION: LIVER SPLEEN
; FILE REFERENCE: 24011-727
; CURRENT APPLICATION NUMBER: US/09/249,697A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/968,800
; PRIOR FILING DATE: 1997-11-22
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 2365
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (205)...(1866)
US-09-249-697A-5

Query Match 1.8%; Score 60.8; DB 4; Length 2365;
Best Local Similarity 90.3%; Pred. No. 9.7e-07;
Matches 65; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy         58 GATTCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTCTGAGGT 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |
Db          7 GATTCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTCTGCAGG 66

Qy        118 GCAAGGACCAGG 129
              |  || ||| ||
Db         67 GACAGCACCCGG 78
```

RESULT 2
US-09-249-697A-18
; Sequence 18, Application US/09249697A
; Patent No. 6392018
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Yeung, George
; TITLE OF INVENTION: NOVEL EGF MOTIF PROTEIN OBTAINED FROM A CDNA LIBRARY OF FETAL
; TITLE OF INVENTION: LIVER SPLEEN
; FILE REFERENCE: 24011-727
; CURRENT APPLICATION NUMBER: US/09/249,697A
; CURRENT FILING DATE: 1999-02-12

```
;  PRIOR APPLICATION NUMBER: US 08/968,800
;  PRIOR FILING DATE: 1997-11-22
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 18
;   LENGTH: 2365
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (205)..(1863)
US-09-249-697A-18

  Query Match             1.8%;  Score 60.8;  DB 4;  Length 2365;
  Best Local Similarity   90.3%;  Pred. No. 9.7e-07;
  Matches   65;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy         58 GATTCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTCTGAGGT 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |
Db          7 GATTCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTCTGCAGG 66

Qy        118 GCAAGGACCAGG 129
              |  || ||| ||
Db         67 GACAGCACCCGG 78


RESULT 3
US-09-363-316B-5
; Sequence 5, Application US/09363316B
; Patent No. 6392019
; GENERAL INFORMATION:
;  APPLICANT: Ford, John
;  APPLICANT:  Yeung, George
;  TITLE OF INVENTION: EGF MOTIF PROTEIN MATERIALS AND METHODS
;  FILE REFERENCE: 28110/35852
;  CURRENT APPLICATION NUMBER: US/09/363,316B
;  CURRENT FILING DATE:  1999-07-28
;  PRIOR APPLICATION NUMBER: US 09/249,697
;  PRIOR FILING DATE: 1999-02-12
;  PRIOR APPLICATION NUMBER: US 08/968,800
;  PRIOR FILING DATE: 1997-11-22
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 2365
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (205)...(1866)
;   NAME/KEY: misc_feature
;   LOCATION: (357)
;   OTHER INFORMATION: Xaa = any amino acid
US-09-363-316B-5

  Query Match             1.8%;  Score 60.8;  DB 4;  Length 2365;
  Best Local Similarity   90.3%;  Pred. No. 9.7e-07;
  Matches   65;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy         58 GATTCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTCTGAGGT 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |
Db          7 GATTCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTCTGCAGG 66

Qy        118 GCAAGGACCAGG 129
              |  || ||| ||
Db         67 GACAGCACCCGG 78


RESULT 4
US-09-363-316B-23
; Sequence 23, Application US/09363316B
; Patent No. 6392019
; GENERAL INFORMATION:
;  APPLICANT: Ford, John
;  APPLICANT:  Yeung, George
;  TITLE OF INVENTION: EGF MOTIF PROTEIN MATERIALS AND METHODS
;  FILE REFERENCE: 28110/35852
;  CURRENT APPLICATION NUMBER: US/09/363,316B
;  CURRENT FILING DATE:  1999-07-28
;  PRIOR APPLICATION NUMBER: US 09/249,697
;  PRIOR FILING DATE: 1999-02-12
;  PRIOR APPLICATION NUMBER: US 08/968,800
;  PRIOR FILING DATE: 1997-11-22
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 23
;   LENGTH: 2365
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (205)..(1863)
US-09-363-316B-23

  Query Match             1.8%;  Score 60.8;  DB 4;  Length 2365;
  Best Local Similarity   90.3%;  Pred. No. 9.7e-07;
  Matches   65;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy         58 GATTCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTCTGAGGT 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |
Db          7 GATTCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTCTGCAGG 66

Qy        118 GCAAGGACCAGG 129
              |  || ||| ||
Db         67 GACAGCACCCGG 78


RESULT 5
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
```

```
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             1.7%;  Score 57.6;  DB 1;  Length 7218;
  Best Local Similarity   3.1%;  Pred. No. 1.5e-05;
  Matches   12;  Conservative  227;  Mismatches  152;  Indels    0;  Gaps    0;

Qy        104 GGCAGGTCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAA 163
              :: :::   :::: : :::::  ::::  :::: ::::: :: ::   :::::: :::::
Db       1434 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1375

Qy        164 ACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCG 223
              : :: ::  :::::::   :::  ::::  :::::::::   : ::  :::::: :: :
Db       1374 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1315

Qy        224 CAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCG 283
               ::  :: :::    :  :  ::  :: ::   ::  : :   ::   :::    ::  :
Db       1314 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1255

Qy        284 CCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACA 343
                :  :: :::::::::  :   ::  : ::: :: ::::::::: :: :::::::: :
Db       1254 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1195

Qy        344 GCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAGAGA 403
              :  :::  : :::::    ::: :   :  :::    :::::  :  :  :: :::::::
Db       1194 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1135

Qy        404 GGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCAT 463
              :::::  :  : :::::     : ::    :: : ::: : :::: ::::   :::  :
Db       1134 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1075

Qy        464 GTTAAAGAATCAAAAACAGGTTCTTCAGGCA 494
              :  :::::|||  || |    ||  |  |||
Db       1074 RRRRRRRRATCGCAAGCTCCCTCGACCTGCA 1044


RESULT 6
US-08-938-830-28
; Sequence 28, Application US/08938830
; Patent No. 6040437
;  GENERAL INFORMATION:
;    APPLICANT:  Lasky, Laurence A.
;    APPLICANT:  Dowbenko, Donald J.
;    TITLE OF INVENTION:  Tyrosine Phosphorylated Cleavage
;    TITLE OF INVENTION:  Furrow-Associated Proteins (PSTPIPs)
;    NUMBER OF SEQUENCES:  73
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/938,830
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/798419
;      FILING DATE:  07-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  P1066P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-3216
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1613 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-938-830-28

  Query Match             1.6%;  Score 57;  DB 3;  Length 1613;
  Best Local Similarity   81.5%;  Pred. No. 8.7e-06;
  Matches   66;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy         46 CCGCGGGAATTCGATTCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGG 105
              | || | | |  | ||||||||||||||||||||||||||||||||||||||||||||
Db         67 CTGCAGAATTCGGCTTCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGG 126

Qy        106 CAGGTCTGAGGTGCAAGGACC 126
              |||||||    | ||  | ||
Db        127 CAGGTCTAGAATTCAGCGGCC 147


RESULT 7
US-09-244-314-1
; Sequence 1, Application US/09244314
; Patent No. 6274362
; GENERAL INFORMATION:
;  APPLICANT: Hodge, Martin R.
;  APPLICANT: Yowe, David
;  TITLE OF INVENTION: RGS-Containing Molecules and Uses Thereof
;  FILE REFERENCE: 5800-19, 035800/174680
;  CURRENT APPLICATION NUMBER: US/09/244,314
;  CURRENT FILING DATE: 1999-02-04
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2217
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (160)..(867)
US-09-244-314-1

  Query Match             1.5%;  Score 52.8;  DB 4;  Length 2217;
  Best Local Similarity   89.1%;  Pred. No. 0.00016;
  Matches   57;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy         51 GGAATTCGATTCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGT 110
              | |  |  |||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAATTCGGCTTCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGT 60

Qy        111 CTGA 114
              | |
Db         61 ATAA 64


RESULT 8
US-09-498-959-1
; Sequence 1, Application US/09498959
; Patent No. 6410240
; GENERAL INFORMATION:
;  APPLICANT: Hodge, Martin R.
;  APPLICANT: Yowe, David
;  TITLE OF INVENTION: RGS-Containing Molecules and Uses
;  TITLE OF INVENTION: Thereof
;  FILE REFERENCE: 5800-19A
;  CURRENT APPLICATION NUMBER: US/09/498,959
```

```
;  CURRENT FILING DATE: 2000-02-04
;  EARLIER APPLICATION NUMBER: 09/244,314
;  EARLIER FILING DATE: 1999-02-04
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 2217
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (160)...(867)
US-09-498-959-1

  Query Match             1.5%;  Score 52.8;  DB 4;  Length 2217;
  Best Local Similarity   89.1%;  Pred. No. 0.00016;
  Matches   57;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy         51 GGAATTCGATTCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGT 110
              | | |  | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAATTCGGCTTCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGT 60

Qy        111 CTGA 114
              | |
Db         61 ATAA 64


RESULT 9
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             1.5%;  Score 52.4;  DB 1;  Length 7218;
  Best Local Similarity   4.9%;  Pred. No. 0.00043;
  Matches   20;  Conservative  220;  Mismatches  166;  Indels    0;  Gaps    0;

Qy       2601 CCTACAAGCTGTGCCACATCTATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTT 2660
              | | | | :: : :: : ::: :: :  ::  ::   : :: ::: : ::::: ::::::
Db       1059 CTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1118

Qy       2661 GCCAGTACGCCCACAAGCTGGCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATC 2720
                ::  : : ::: :   ::  :::::::: ::  ::    : ::: :      ::   ::
Db       1119 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1178

Qy       2721 TGTCACTGTCAAACCGCCTTTACTACCTCTAACCTGCAGAAGACGATGCAGCCGCTTTTC 2780
              : :: :: ::   :: ::::: :: ::::::  ::: :      :  : :  :: ::::::
Db       1179 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1238

Qy       2781 TTTTTGAAATGACTTTGGGATTTTTTTAAGCTTTTATTTACTTTTTTTTTTAACTGTTATC 2840
              :::::    :  ::::    :::::::   ::::: ::: :::::::::  :: :: ::
Db       1239 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1298

Qy       2841 TTTCTGGATGAAACTTGGGAAGGGGATTAGGAGATCTAGCATTTTATTTCTAGCATTGCT 2900
              :::::   :   :::          ::     :::  : :::: :::::  : :: ::
Db       1299 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1358

Qy       2901 ATTCACCGGCTTCCTTATTTTATATGTAAAAATTAAGATTTTATATTTTATCTTCTTGTT 2960
              ::: ::  ::::::: :::: : : :      ::    :::: : :::: ::::::: ::
Db       1359 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1418

Qy       2961 TCTCATAGATATTTTGTGAGCATTTTTTTGTTTATTTTGAAGAAAT 3006
              :::: :   : ::::   | ||  || || | | | || ||  | |
Db       1419 YYYYYYYYYYYYYYYYGTACCAAATTCTTCTATCTCTTTAACTACT 1464


RESULT 10
US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827
;  GENERAL INFORMATION:
;    APPLICANT:  Sim, Kim L.
;    APPLICANT:  Chitnis, Chetan
;    APPLICANT:  Miller, Louis H.
;    APPLICANT:  Peterson, David S.
;    APPLICANT:  Su, Xin-zhaun
;    APPLICANT:  Wellems, Thomas E.
;    TITLE OF INVENTION:  BINDING DOMAINS FROM PLASMODIUM VIVAX
;    TITLE OF INVENTION:  AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe Martens Olson & Bear
;      STREET:  620 Newport Center Drive  16th Floor
;      CITY:  Newport Beach
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,826B
;      FILING DATE:  10-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  NIH121.001CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
```

```
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19124 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-487-826B-13

  Query Match             1.5%;  Score 51.8;  DB 2;  Length 19124;
  Best Local Similarity   47.0%;  Pred. No. 0.0011;
  Matches  159;  Conservative    0;  Mismatches  179;  Indels    0;  Gaps    0;

Qy       2776 TTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTTTATTTACTTTTTTTTTTAACTG 2835
             ||||||||||     ||  |||  ||| | | |    |||| |||| | |||||||   |
Db      15720 TTTTCTTTTTTTTTTGTTTTTATGATATATATTTTTTTTTTTTTTAATGTTTTTTTTTTC 15661

Qy       2836 TTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGATCTAGCATTTTATTTCTAGCA 2895
             || | ||  |          ||   ||      ||             | ||| ||| ||
Db      15660 TTCTTTTGTTTTTATTTTTTTTATAATCATTTTTTTTTTTATATAAAATTTTTTTTTAATT 15601

Qy       2896 TTGCTATTCACCGGCTTCCTTATTTTATATGTAAAAATTAAGATTTTATATTTTATCTTC 2955
             ||  | |         ||     ||||   | ||  ||      || || |||| ||  ||
Db      15600 TTTTTTTGATAATCTTTTTCATTTTTTATTCTATCAAAATTTATATTTTATTATAATTTT 15541

Qy       2956 TTGTTTCTCATAGATATTTTGTGAGCATTTTTTTTGTTTATTTTGAAGAAATGTGGATAAG 3015
             |  | | |  || | | ||| |   | ||||||| ||| ||||    |  |    | |
Db      15540 TATTATTTTTTAAAAAATTTTTCTCCTTTTTTTTTTTTTTTTTTTTTTTATTTTAATAAATT 15481

Qy       3016 ATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTGAAATGTGTTTGGAGATTTACT 3075
              |  ||  || |  |        | | | |   || |  |  ||| |||  |  | |  |
Db      15480 TTTTTTTATATTTCATTTTTCTTTTTTTCATTTTAATAAATTTGTTTTTATATTTCTTTT 15421

Qy       3076 TAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAA 3113
             | ||  ||  | ||    |    ||    ||  || ||
Db      15420 TTAATATAAATACATATATATAAAATATATATATAAAA 15383


RESULT 11
US-08-592-126-73/c
; Sequence 73, Application US/08592126
; Patent No. 5821091
;  GENERAL INFORMATION:
;    APPLICANT:  Gregory Dolganov
;    TITLE OF INVENTION:  Transcripts Encoding Immunomodulatory
;    TITLE OF INVENTION:  Polypeptides
;    NUMBER OF SEQUENCES:  151
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/592,126
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sholtz, Charles K.
;      REGISTRATION NUMBER:  38,615
;      REFERENCE/DOCKET NUMBER:  4600-0111
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  73:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  797 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  G248b.seq
US-08-592-126-73

  Query Match             1.5%;  Score 51;  DB 1;  Length 797;
  Best Local Similarity   91.5%;  Pred. No. 0.00027;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         51 GGAATTCGATTCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGG 109
              | | |  | ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        797 GAATTCGGCTTCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGG 739


RESULT 12
US-09-141-206-1
; Sequence 1, Application US/09141206
; Patent No. 6187559
; GENERAL INFORMATION:
;  APPLICANT: Steed, Paul M.
;  APPLICANT: Lasala, Daniel J.
;  TITLE OF INVENTION: Amino Acid Sequence of Human PLD2A
; Patent No. 6187559
;  FILE REFERENCE: 4-30148/P1/CGC1954/R
;  CURRENT APPLICATION NUMBER: US/09/141,206
;  CURRENT FILING DATE: 1998-08-27
;  EARLIER APPLICATION NUMBER: 60/057,802
;  EARLIER FILING DATE: 1997-08-28
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 3388
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (162)..(2963)
;   OTHER INFORMATION: Human PLD2
US-09-141-206-1

  Query Match             1.5%;  Score 50.4;  DB 4;  Length 3388;
  Best Local Similarity   98.1%;  Pred. No. 0.00096;
  Matches   51;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         62 CCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTCTG 113
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db          1 CCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTCCG 52


RESULT 13
US-08-866-757-1
; Sequence 1, Application US/08866757
; Patent No. 5858716
;  GENERAL INFORMATION:
;    APPLICANT:  ELSHOURBAGY, NABIL A
;    APPLICANT:  LI, XIAOTONG
;    APPLICANT:  BERGSMA, DERK J
;    TITLE OF INVENTION:  NOVEL 7TM RECEPTOR (H2CAA71)
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RATNER & PRESTIA
;      STREET:  P.O. BOX 980
;      CITY:  VALLEY FORGE
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/866,757
;      FILING DATE:  30-MAY-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PRESTIA, PAUL F
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  GH-70055
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0701
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4203 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-866-757-1

  Query Match             1.4%;  Score 50.2;  DB 2;  Length 4203;
  Best Local Similarity   68.0%;  Pred. No. 0.0012;
  Matches   70;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         58 GATTCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTCTGAGGT 117
              | ||||||||||||||| |||||||||||||||||||||||||||||||||||    |
Db          1 GCTTCCATCCTAATACAACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTGCTTGAC 60

Qy        118 GCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAG 160
              | | |  || |  |     |   |  | | | |||    | ||
Db         61 GGAGGTGCCTGTGCACCCCCTCAGCAATCTGCCCACCCTACAG 103


RESULT 14
US-09-153-593-1
; Sequence 1, Application US/09153593A
; Patent No. 6174994
; GENERAL INFORMATION:
;  APPLICANT: ELSHOURBAGY, NABIL A
;  APPLICANT: LI, XIAOTONG
;  APPLICANT: BERGSMA, DERK J
;  TITLE OF INVENTION: NOVEL 7TM RECEPTOR (H2CAA71)
;  FILE REFERENCE: GH-70055-1
;  CURRENT APPLICATION NUMBER: US/09/153,593A
;  CURRENT FILING DATE: 1998-09-15
;  EARLIER APPLICATION NUMBER: 08/866,757
;  EARLIER FILING DATE: 1997-05-30
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 4203
;   TYPE: DNA
;   ORGANISM: HOMO SAPIENS
US-09-153-593-1

  Query Match             1.4%;  Score 50.2;  DB 4;  Length 4203;
  Best Local Similarity   68.0%;  Pred. No. 0.0012;
  Matches   70;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         58 GATTCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTCTGAGGT 117
              | ||||||||||||||| |||||||||||||||||||||||||||||||||||    |
Db          1 GCTTCCATCCTAATACAACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTGCTTGAC 60

Qy        118 GCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAG 160
              | | |  || |  |     |   |  | | | |||    | ||
Db         61 GGAGGTGCCTGTGCACCCCCTCAGCAATCTGCCCACCCTACAG 103


RESULT 15
US-08-998-416-595
; Sequence 595, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  595:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  658 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1408RP
US-08-998-416-595

  Query Match             1.4%;  Score 50;  DB 4;  Length 658;
  Best Local Similarity   50.0%;  Pred. No. 0.00045;
  Matches  151;  Conservative    0;  Mismatches  150;  Indels    1;  Gaps    1;

Qy       2776 TTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTTTATTTACTTTTTTTTTAACTG 2835
              | || ||||   | | | |||   |||| ||| |   |||||||  || ||| ||   |
Db        275 TATTATTTTATTATTTATTTTATTATTTATTTTATTATTTATTTTATTATTTATTTTATT 334

Qy       2836 TTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGATCTAGCATTTTATTTCTAGCA 2895
              | | ||| |  || |    |       |    | |     ||   |||||||| | |
Db        335 ATTTATTTATTTATTATTTATTTATTTGTTTGTTTGTTTATTATTTTTTTATTTATTACC 394

Qy       2896 TTGCTATTCACCGGCTTCCTTATTTTATATGTAAAAATTAAGATTTTATATTTTATCTTC 2955
              ||  ||||    |   | |  |||  | | |    | | ||  | || || |||| ||
Db        395 TTTTTATTTTATGTTATTTTATTTTATTTTATTTTTACTTAGTATATA-ATATTATATTA 453

Qy       2956 TTGTTTCTCATAGATATTTTGTGAGCATTTTTTTGTTTATTTTGAAGAAATGTGGATAAG 3015
```

```
Db        454 TATCATAGTATAGTTATATTATGGTGACTTTATTCATTATATAGATTGTATTTTGTGAAC 513

Qy       3016 ATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTGAAATGTGTTTGGAGATTTACT 3075

Db        514 ATAATATATATGCTATTTCTATTTCTATTTTATTTTATTTTATTTTATTTTATTTTATTT 573

Qy       3076 TA 3077

Db        574 TA 575
```

Search completed: July 22, 2003, 06:25:09
Job time : 142.728 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2003, 18:26:42 ; Search time 670.254 Seconds
(without alignments)
10686.606 Million cell updates/sec

Title: US-09-873-737A-5
Perfect score: 3472
Sequence: 1 gggcgaattgggcccgacgt.....taaaataaaattgatgtccc 3472

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
11: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq3:*
13: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	3469	99.9	3472	10	US-09-873-737A-5	Sequence 5, Appli
	2	2311	66.6	2328	14	US-10-043-774B-1	Sequence 1, Appli
	3	2291	66.0	2292	11	US-09-998-598-2602	Sequence 2602, Ap
	4	1833.6	52.8	4064	10	US-09-873-737A-3	Sequence 3, Appli
c	5	552.2	15.9	582	11	US-09-998-598-2600	Sequence 2600, Ap
c	6	550.4	15.9	590	11	US-09-919-580-634	Sequence 634, App
	7	459.4	13.2	501	11	US-09-920-300A-34	Sequence 34, Appl
	8	459.4	13.2	501	15	US-10-033-528-34	Sequence 34, Appl
c	9	458.4	13.2	460	10	US-09-777-564-1706	Sequence 1706, Ap
c	10	458.4	13.2	460	15	US-10-015-219-1706	Sequence 1706, Ap
	11	457.4	13.2	2009	15	US-10-029-495-8	Sequence 8, Appli
	12	395.4	11.4	423	15	US-10-102-806-267	Sequence 267, App
	13	377	10.9	3649	10	US-09-801-574-15	Sequence 15, Appl
	14	253	7.3	3047	10	US-09-873-737A-1	Sequence 1, Appli
	15	192	5.5	560	15	US-10-029-495-1	Sequence 1, Appli
	16	188.2	5.4	889	15	US-10-027-632-166243	Sequence 166243,
	17	188.2	5.4	889	15	US-10-027-632-166244	Sequence 166244,
	18	188.2	5.4	889	15	US-10-027-632-166245	Sequence 166245,
	19	111.8	3.2	2914	11	US-09-774-434-2	Sequence 2, Appli
	20	95.6	2.8	444	10	US-09-815-343-3	Sequence 3, Appli
	21	88.8	2.6	7478	15	US-10-007-078-3	Sequence 3, Appli
	22	84.6	2.4	1238	15	US-10-174-363-31	Sequence 31, Appl
	23	84	2.4	677	15	US-10-027-632-140400	Sequence 140400,
	24	84	2.4	677	15	US-10-027-632-140401	Sequence 140401,
	25	83.6	2.4	490	12	US-09-918-995-10623	Sequence 10623, A
	26	81.8	2.4	3151	15	US-10-174-363-41	Sequence 41, Appl
	27	78.6	2.3	213	10	US-09-864-761-32092	Sequence 32092, A
	28	78.6	2.3	1619	15	US-10-174-363-25	Sequence 25, Appl
	29	76.6	2.2	3096	15	US-10-174-363-7	Sequence 7, Appli
	30	74.2	2.1	441	12	US-09-918-995-15958	Sequence 15958, A
	31	70.8	2.0	2827	15	US-10-174-363-3	Sequence 3, Appli
	32	70.4	2.0	1565	15	US-10-174-363-49	Sequence 49, Appl
	33	70.2	2.0	3549	15	US-10-174-363-27	Sequence 27, Appl
	34	70.2	2.0	3705	15	US-10-174-363-53	Sequence 53, Appl
	35	68	2.0	364	10	US-09-864-761-4964	Sequence 4964, Ap
	36	67.6	1.9	2042	15	US-10-174-363-1	Sequence 1, Appli
	37	67	1.9	446	11	US-09-960-352-3400	Sequence 3400, Ap
	38	66.6	1.9	3239	15	US-10-174-363-39	Sequence 39, Appl
	39	64.8	1.9	2793	11	US-09-938-842A-394	Sequence 394, App
	40	64.6	1.9	3714	15	US-10-174-363-13	Sequence 13, Appl
	41	64.6	1.9	3808	15	US-10-174-363-11	Sequence 11, Appl
	42	64.2	1.8	166	10	US-09-864-761-21693	Sequence 21693, A
c	43	62.8	1.8	288	15	US-10-137-036-38	Sequence 38, Appl
	44	62.2	1.8	3009	15	US-10-174-363-21	Sequence 21, Appl
	45	62	1.8	3357	15	US-10-161-403-96	Sequence 96, Appl

ALIGNMENTS

RESULT 1
US-09-873-737A-5
; Sequence 5, Application US/09873737A
; Patent No. US20020076797A1
; GENERAL INFORMATION:
; APPLICANT: Duke University
; APPLICANT: Lin, Haifan
; TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
; TITLE OF INVENTION: PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
; FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2
; CURRENT APPLICATION NUMBER: US/09/873,737A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/US99/28764
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/110,901
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 3472
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (167)..(2749)
; NAME/KEY: misc_feature
; LOCATION: (392)
; OTHER INFORMATION: n=t or a, Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (1073)
; OTHER INFORMATION: n=c or a, Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (2369)
; OTHER INFORMATION: n=c or a, Xaa=Leu or Ile
US-09-873-737A-5

Query Match 99.9%; Score 3469; DB 10; Length 3472;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GGGCGAATTGGGCCCGACGTCGCATGCTCCCGGCCGCCATGGCGGCCGCGGGAATTCGAT 60
Db       1 GGGCGAATTGGGCCCGACGTCGCATGCTCCCGGCCGCCATGGCGGCCGCGGGAATTCGAT 60

Qy      61 TCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTCTGAGGTGCA 120
Db      61 TCCATCCTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTCTGAGGTGCA 120

Qy     121 AGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAACAATGACTGGGAGAGC 180
Db     121 AGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAACAATGACTGGGAGAGC 180

Qy     181 CCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGCTGGTGGGCTCCAC 240
Db     181 CCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGCTGGTGGGCTCCAC 240

Qy     241 TGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGCCACCAGCAGAGGGGGA 300
Db     241 TGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGCCACCAGCAGAGGGGGA 300

Qy     301 ATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAGCCAAGTCACAAGGACT 360
Db     301 ATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAGCCAAGTCACAAGGACT 360

Qy     361 CCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAGAGAGGAGGTCGTCGTAGAGA 420
Db     361 CCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAGAGAGGAGGTCGTCGTAGAGA 420

Qy     421 TTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGTTAAAGAATCAAAAAC 480
Db     421 TTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGTTAAAGAATCAAAAAC 480

Qy     481 AGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGACATCCCGTCC 540
Db     481 AGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGACATCCCGTCC 540

Qy     541 CCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGGAAGCCAGAAGACT 600
Db     541 CCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGGAAGCCAGAAGACT 600

Qy     601 CCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCATGCTTTTGATGG 660
Db     601 CCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCATGCTTTTGATGG 660

Qy     661 AACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTTTTTAGTAAGAC 720
Db     661 AACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTTTTTAGTAAGAC 720

Qy     721 CCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATGAACTTCCACCTACATC 780
Db     721 CCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATGAACTTCCACCTACATC 780

Qy     781 ACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAAATCATGAATTT 840
Db     781 ACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAAATCATGAATTT 840

Qy     841 GCAACAAATTGGACGAAATTATTATAACCCAAATGACCCAATTGATATTCCAAGTCACAG 900
Db     841 GCAACAAATTGGACGAAATTATTATAACCCAAATGACCCAATTGATATTCCAAGTCACAG 900

Qy     901 GTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAACAGCATCATGCT 960
Db     901 GTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAACAGCATCATGCT 960

Qy     961 CTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGATTTCATGTTCAA 1020
Db     961 CTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGATTTCATGTTCAA 1020

Qy    1021 CTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAAGAANTAATAGG 1080
Db    1021 CTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAAGAANTAATAGG 1080

Qy    1081 TTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGAGTGGATGATATTGACTGGGA 1140
Db    1081 TTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGAGTGGATGATATTGACTGGGA 1140

Qy    1141 CCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGGGGTCAGCTTCTTAGAATA 1200
Db    1141 CCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCTGGGGTCAGCTTCTTAGAATA 1200

Qy    1201 CTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGCCA 1260
Db    1201 CTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGCCA 1260

Qy    1261 GCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATGCTCATTCCTGA 1320
Db    1261 GCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATGCTCATTCCTGA 1320

Qy    1321 GCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAACGTGATGAAAGA 1380
Db    1321 GCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAACGTGATGAAAGA 1380

Qy    1381 CTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTGGGACGACTCAT 1440
Db    1381 CTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTGGGACGACTCAT 1440

Qy    1441 TGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGGGTTTGAGCTT 1500
Db    1441 TGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGGGTTTGAGCTT 1500

Qy    1501 TGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAGATTCACCAAGG 1560
Db    1501 TGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAGATTCACCAAGG 1560

Qy    1561 TGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAACAAGAGGTGC 1620
Db    1561 TGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAACAAGAGGTGC 1620

Qy    1621 ACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACGCGAAGAAATTA 1680
Db    1621 ACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACGCGAAGAAATTA 1680

Qy    1681 TGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGCATGCA 1740
Db    1681 TGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGCATGCA 1740

Qy    1741 AATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACTGAAGCCTACTTAAGAGTCTT 1800
Db    1741 AATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACTGAAGCCTACTTAAGAGTCTT 1800

Qy    1801 ACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCTGTTGTCAAGTAATCGGAA 1860
Db    1801 ACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCTGTTGTCAAGTAATCGGAA 1860

Qy    1861 GGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAGTG 1920
Db    1861 GGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAGTG 1920

Qy    1921 TGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAAAGATTGC 1980
Db    1921 TGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAAAGATTGC 1980

Qy    1981 CCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCT 2040
Db    1981 CCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCT 2040

Qy    2041 CGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGC 2100
Db    2041 CGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGC 2100

Qy    2101 AGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCATATTTCA 2160
Db    2101 AGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCATATTTCA 2160

Qy    2161 GGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGGCTCTGAGGGC 2220
```

```
Db       2161 GGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGGCTCTGAGGGC 2220

Qy       2221 TTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATCGTGTACCGCGATGGCGTAGG 2280

Db       2221 TTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATCGTGTACCGCGATGGCGTAGG 2280

Qy       2281 AGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCACAGTTTTTGGATTGTCTAAA 2340

Db       2281 AGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCACAGTTTTTGGATTGTCTAAA 2340

Qy       2341 ATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATTGTGGTGAAGAAAAGAGTGAA 2400

Db       2341 ATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATTGTGGTGAAGAAAAGAGTGAA 2400

Qy       2401 CACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTCCTGGAACAGTTAT 2460

Db       2401 CACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTCCTGGAACAGTTAT 2460

Qy       2461 TGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTTATCGTGAGCCAGGCTGTGAG 2520

Db       2461 TGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTTATCGTGAGCCAGGCTGTGAG 2520

Qy       2521 AAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTATGACAACAGCGGCCTGAAGCC 2580

Db       2521 AAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTATGACAACAGCGGCCTGAAGCC 2580

Qy       2581 AGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATCTATTACAACTGGCCAGGTGT 2640

Db       2581 AGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATCTATTACAACTGGCCAGGTGT 2640

Qy       2641 CATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTGTTGGCCAGAG 2700

Db       2641 CATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTGTTGGCCAGAG 2700

Qy       2701 TATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAACCTGCAGA 2760

Db       2701 TATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAACCTGCAGA 2760

Qy       2761 AGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTTTATTTA 2820

Db       2761 AGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTTTATTTA 2820

Qy       2821 CTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGATCTAGC 2880

Db       2821 CTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGATCTAGC 2880

Qy       2881 ATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATATGTAAAAATTAAGATT 2940

Db       2881 ATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATATGTAAAAATTAAGATT 2940

Qy       2941 TTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTGTTTATTTTGA 3000

Db       2941 TTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTGTTTATTTTGA 3000

Qy       3001 AGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTGAAATGT 3060

Db       3001 AGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTGAAATGT 3060

Qy       3061 GTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAAACCTATA 3120

Db       3061 GTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAAACCTATA 3120

Qy       3121 TTAACTTTATTTTTGAGATACCTGTTTTGAATTTAAAGGAGATAAGAGGCGTAAAGTAGG 3180

Db       3121 TTAACTTTATTTTTGAGATACCTGTTTTGAATTTAAAGGAGATAAGAGGCGTAAAGTAGG 3180

Qy       3181 ATGCTCACTACAACCATAGGTGGGGTTTCAGCTCATATCTTAAAGATAAAAGGTACTATT 3240

Db       3181 ATGCTCACTACAACCATAGGTGGGGTTTCAGCTCATATCTTAAAGATAAAAGGTACTATT 3240

Qy       3241 ATATAACCTATACACAAGATACAGGAGAAAATATGCTTGATTTTTATTTGGCAGGGGGGC 3300

Db       3241 ATATAACCTATACACAAGATACAGGAGAAAATATGCTTGATTTTTATTTGGCAGGGGGGC 3300

Qy       3301 TAGGTTGTATGGGAGTAAAAAAAACATTGAAAATTTTTAAATTGTCCAAAGAAACATTTT 3360

Db       3301 TAGGTTGTATGGGAGTAAAAAAAACATTGAAAATTTTTAAATTGTCCAAAGAAACATTTT 3360

Qy       3361 AAGACTCTTTAACAAAAAAGGCCATGAGTAAATCTCTATATTAACATCACTATTTATTTT 3420

Db       3361 AAGACTCTTTAACAAAAAAGGCCATGAGTAAATCTCTATATTAACATCACTATTTATTTT 3420

Qy       3421 GTTTTGGAACTGGGACATGATTCTATTTGTTATAAAATAAAATTGATGTCCC 3472

Db       3421 GTTTTGGAACTGGGACATGATTCTATTTGTTATAAAATAAAATTGATGTCCC 3472
```

RESULT 2
US-10-043-774B-1
; Sequence 1, Application US/10043774B
; Publication No. US20030129599A1
; GENERAL INFORMATION:
; APPLICANT: University of Illinois at Chicago
; APPLICANT: Sharma, Arun
; APPLICANT: Hoffman, Ronald
; TITLE OF INVENTION: HUMAN HEMATOPOIETIC GROWTH REGULATORY GENE AND USES
; FILE REFERENCE: MBHB: CU08/PPA
; CURRENT APPLICATION NUMBER: US/10/043,774B
; CURRENT FILING DATE: 2002-06-06
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 2328
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2328)
; OTHER INFORMATION: Human Hiwi Protein
US-10-043-774B-1

Query Match 66.6%; Score 2311; DB 14; Length 2328;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2314; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy        432 TTGGTGTGAATACAAGGCAGAACCTAGACCATGTTAAAGAATCAAAAACAGGTTCTTCAG 491

Db          8 TTGGTGTGAACACAAGGCAGAACTTAGACCATGTTAAAGAATCAAAAACAGGTTCTTCAG 67

Qy        492 GCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGACATCCCGTCCCCAGTGGGCCT 551

Db         68 GCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGACATCCCGTCCCCAGTGGGCCT 127

Qy        552 TATATCAGTATCACATTGACTATAACCCACTGATGGAAGCCAGAAGACTCCGTTCAGCTC 611

Db        128 TATATCAGTATCACATTGACTATAACCCACTGATGGAAGCCAGAAGACTCCGTTCAGCTC 187

Qy        612 TTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCATGCTTTTGATGGAACGATATTAT 671

Db        188 TTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCATGCTTTTGATGGAACGATATTAT 247

Qy        672 TTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTTTTTAGTAAGACCCGGAATGGAG 731

Db        248 TTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTTTTTAGTAAGACCCGGAATGGAG 307

Qy        732 AGGATGTGAGGATAACGATCACTTTAACAAATGAACTTCCACCTACATCACCAACTTGTT 791

Db        308 AGGATGTGAGGATAACGATCACTTTAACAAATGAACTTCCACCTACATCACCAACTTGTT 367

Qy        792 TGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAAATCATGAATTTGCAACAAATTG 851

Db        368 TGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAAATCATGAATTTGCAACAAATTG 427

Qy        852 GACGAAATTATTATAACCCAAATGACCCAATTGATATTCCAAGTCACAGGTTGGTGATTT 911
```

```
Db        428 GACGAAATTATTATAACCCAAATGACCCAATTGATATTCCAAGTCACAGGTTGGTGATTT 487

Qy        912 GGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAACAGCATCATGCTCTGCACTGACG 971
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        488 GGCCTGGCTTCACTACTTCCATCCTTCAGTATGAAAACAGCATCATGCTCTGCACTGACG 547

Qy        972 TTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGATTTCATGTTCAACTTTTATCATC 1031
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        548 TTAGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGATTTCATGTTCAACTTTTATCATC 607

Qy       1032 AGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAAGAANTAATAGGTTTAGTTGTTC 1091
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        608 AGACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAAGAACTAATAGGTTTAGTTGTTC 667

Qy       1092 TTACCAAGTATAACATTAAGACATACAGAGTGGATGATATTGACTGGGACCAGAATCCCA 1151
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        668 TTACCAAGTATAACAATAAGACATACAGAGTGGATGATATTGACTGGGACCAGAATCCCA 727

Qy       1152 AGAGCACCTTTAAGAAAGCCGACGGCTCTGGGGTCAGCTTCTTAGAATACTACAGGAAGC 1211
              |||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db        728 AGAGCACCTTTAAGAAAGCCGACGGCTCTGAAGTCAGCTTCTTAGAATACTACAGGAAGC 787

Qy       1212 AATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGCCAGCCCAAGAGAA 1271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        788 AATACAACCAAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGCCAGCCCAAGAGAA 847

Qy       1272 GGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATGCTCATTCCTGAGCTCTGCTATC 1331
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 GGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATGCTCATTCCTGAGCTCTGCTATC 907

Qy       1332 TTACAGGTCTAACTGATAAAATGCGTAATGATTTTAACGTGATGAAAGACTTAGCCGTTC 1391
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 TTACAGGTCTAACTGATAAAATGCGTAATGATTTTAACGTGATGAAAGACTTAGCCGTTC 967

Qy       1392 ATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTGGGACGACTCATTGATTACATTC 1451
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 ATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTGGGACGACTCATTGATTACATTC 1027

Qy       1452 ATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGGGTTTGAGCTTTGATTCCAACT 1511
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 ATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGGGTTTGAGCTTTGATTCCAACT 1087

Qy       1512 TACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAGATTCACCAAGGTGGAAAAACAT 1571
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 TACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAGATTCACCAAGGTGGAAAAACAT 1147

Qy       1572 TTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAACAAGAGGTGCACCATTAATTA 1631
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 TTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAACAAGAGGTGCACCATTAATTA 1207

Qy       1632 GTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACGCGAAGAAATTATGAAGCAGCCA 1691
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 GTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACGCGAAGAAATTATGAAGCAGCCA 1267

Qy       1692 ATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGCATGCAAATGAGAAAAG 1751
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1268 ATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGCATGCAAATGAGAAAAG 1327

Qy       1752 CAATAATGATTGAAGTGGATGACAGAACTGAAGCCTACTTAAGAGTCTTACAGCAAAAGG 1811
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1328 CAATAATGATTGAAGTGGATGACAGAACTGAAGCCTACTTAAGAGTCTTACAGCAAAAGG 1387

Qy       1812 TCACAGCAGACACCCAGATAGTTGTCTGTCTGTTGTCAAGTAATCGGAAGGACAAATACG 1871
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1388 TCACAGCAGACACCCAGATAGTTGTCTGTCTGTTGTCAAGTAATCGGAAGGACAAATACG 1447

Qy       1872 ATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCC 1931
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1448 ATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCC 1507

Qy       1932 GAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAAAGATTGCCCTACAGATGA 1991
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1508 GAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAAAGATTGCCCTACAGATGA 1567

Qy       1992 ACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCTCGTGATGATCG 2051
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1568 ACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCTCGTGATGATCG 1627

Qy       2052 TTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTG 2111
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1628 TTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTG 1687

Qy       2112 CCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCATATTTCAGGATAGAGGAC 2171
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1688 CCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCATATTTCAGGATAGAGGAC 1747

Qy       2172 AGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGGCTCTGAGGGCTTGGAATAGCT 2231
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1748 AGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGGCTCTGAGGGCTTGGAATAGCT 1807

Qy       2232 GCAATGAGTACATGCCCAGCCGGATCATCGTGTACCGCGATGGCGTAGGAGACGGCCAGC 2291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1808 GCAATGAGTACATGCCCAGCCGGATCATCGTGTACCGCGATGGCGTAGGAGACGGCCAGC 1867

Qy       2292 TGAAAACACTGGTGAACTACGAAGTGCCACAGTTTTTGGATTGTCTAAAATCCATTGGTA 2351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1868 TGAAAACACTGGTGAACTACGAAGTGCCACAGTTTTTGGATTGTCTAAAATCCATTGGTA 1927

Qy       2352 GAGGTTACAACCCTAGANTAACGGTAATTGTGGTGAAGAAAAGAGTGAACACCAGATTTT 2411
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       1928 GAGGTTACAACCCTAGACTAACGGTAATTGTGGTGAAGAAAAGAGTGAACACCAGATTTT 1987

Qy       2412 TTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTCCTGGAACAGTTATTGATGTAGAGG 2471
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1988 TTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTCCTGGAACAGTTATTGATGTAGAGG 2047

Qy       2472 TTACCAGACCAGAATGGTATGACTTTTTTATCGTGAGCCAGGCTGTGAGAAGTGGTAGTG 2531
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2048 TTACCAGACCAGAATGGTATGACTTTTTTATCGTGAGCCAGGCTGTGAGAAGTGGTAGTG 2107

Qy       2532 TTTCTCCCACACATTACAATGTCATCTATGACAACAGCGGCCTGAAGCCAGACCACATAC 2591
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2108 TTTCTCCCACACATTACAATGTCATCTATGACAACAGCGGCCTGAAGCCAGACCACATAC 2167

Qy       2592 AGCGCTTGACCTACAAGCTGTGCCACATCTATTACAACTGGCCAGGTGTCATTCGTGTTC 2651
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2168 AGCGCTTGACCTACAAGCTGTGCCACATCTATTACAACTGGCCAGGTGTCATTCGTGTTC 2227

Qy       2652 CTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTGTTGGCCAGAGTATTCACAGAG 2711
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2228 CTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTGTTGGCCAGAGTATTCACAGAG 2287

Qy       2712 AGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAA 2752
              |||||||||||||||||||||||||||||||||||||||||
Db       2288 AGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAA 2328
```

```
RESULT 3
US-09-998-598-2602
; Sequence 2602, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
;  APPLICANT: Stolk, John A.
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Chenault, Ruth A.
;  APPLICANT:  Meagher, Madelein Joy
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.561
;  CURRENT APPLICATION NUMBER: US/09/998,598
;  CURRENT FILING DATE:  2001-11-16
;  NUMBER OF SEQ ID NOS: 2606
;  SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 2602
;   LENGTH: 2292
;   TYPE: DNA
```

; ORGANISM: Homo sapiens
US-09-998-598-2602

Query Match 66.0%; Score 2291; DB 11; Length 2292;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2291; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1181 GGGGTCAGCTTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAG 1240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGGGTCAGCTTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAG 60

Qy       1241 CAGCCTGTCTTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGG 1300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGCCTGTCTTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGG 120

Qy       1301 CCTGCCATGCTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAAT 1360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTGCCATGCTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAAT 180

Qy       1361 GATTTTAACGTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAG 1420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTTTAACGTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAG 240

Qy       1421 CGTGAAGTGGGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTT 1480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CGTGAAGTGGGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTT 300

Qy       1481 CGAGACTGGGGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAA 1540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGAGACTGGGGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAA 360

Qy       1541 ACAGAAAAGATTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGG 1600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACAGAAAAGATTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGG 420

Qy       1601 TCCAAAGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTG 1660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCAAAGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTG 480

Qy       1661 ATCTATACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTT 1720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATCTATACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTT 540

Qy       1721 ACACCAGCCATGGGCATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACT 1780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACACCAGCCATGGGCATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACT 600

Qy       1781 GAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGT 1840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGT 660

Qy       1841 CTGTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGAT 1900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CTGTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGAT 720

Qy       1901 TGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATG 1960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATG 780

Qy       1961 GCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGG 2020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGG 840

Qy       2021 GTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACA 2080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACA 900

Qy       2081 GCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGG 2140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGG 960

Qy       2141 TTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGC 2200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 TTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGC 1020

Qy       2201 CTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATC 2260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 CTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATC 1080

Qy       2261 GTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCA 2320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 GTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCA 1140

Qy       2321 CAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATT 2380
             ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        1141 CAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGACTAACGGTAATT 1200

Qy       2381 GTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAAT 2440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 GTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAAT 1260

Qy       2441 CCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTT 2500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1261 CCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTT 1320

Qy       2501 ATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTAT 2560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1321 ATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTAT 1380

Qy       2561 GACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATC 2620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1381 GACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATC 1440

Qy       2621 TATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTG 2680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1441 TATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTG 1500

Qy       2681 GCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTT 2740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1501 GCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTT 1560

Qy       2741 TACTACCTCTAACCTGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGA 2800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1561 TACTACCTCTAACCTGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGA 1620

Qy       2801 TTTTTTTAAGCTTTTATTTACTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGA 2860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1621 TTTTTTTAAGCTTTTATTTACTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGA 1680

Qy       2861 AGGGGATTAGGAGATCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTT 2920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1681 AGGGGATTAGGAGATCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTT 1740

Qy       2921 TATATGTAAAAATTAAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAG 2980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1741 TATATGTAAAAATTAAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAG 1800

Qy       2981 CATTTTTTTGTTTATTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTC 3040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1801 CATTTTTTTGTTTATTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTC 1860

Qy       3041 TCTGAGAGTATTTGAAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAA 3100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1861 TCTGAGAGTATTTGAAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAA 1920

Qy       3101 GTCCTACTTATAAACCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTTAAAGGA 3160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1921 GTCCTACTTATAAACCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTTAAAGGA 1980

Qy       3161 GATAAGAGGCGTAAAGTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTCATATCT 3220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1981 GATAAGAGGCGTAAAGTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTCATATCT 2040

Qy       3221 TAAAGATAAAAGGTACTATTATATAACCTATACACAAGATACAGGAGAAAATATGCTTGA 3280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2041 TAAAGATAAAAGGTACTATTATATAACCTATACACAAGATACAGGAGAAAATATGCTTGA 2100
```

```
Qy       3281 TTTTTATTTGGCAGGGGGGCTAGGTTGTATGGGAGTAAAAAAAACATTGAAAATTTTTAA 3340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TTTTTATTTGGCAGGGGGGCTAGGTTGTATGGGAGTAAAAAAAACATTGAAAATTTTTAA 2160

Qy       3341 ATTGTCCAAAGAAACATTTTAAGACTCTTTAACAAAAAAGGCCATGAGTAAATCTCTATA 3400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ATTGTCCAAAGAAACATTTTAAGACTCTTTAACAAAAAAGGCCATGAGTAAATCTCTATA 2220

Qy       3401 TTAACATCACTATTTATTTTGTTTTGGAACTGGGACATGATTCTATTTGTTATAAAATAA 3460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 TTAACATCACTATTTATTTTGTTTTGGAACTGGGACATGATTCTATTTGTTATAAAATAA 2280

Qy       3461 AATTGATGTCCC 3472
             ||||||||||||
Db       2281 AATTGATGTCCC 2292


RESULT 4
US-09-873-737A-3
; Sequence 3, Application US/09873737A
; Patent No. US20020076797A1
; GENERAL INFORMATION:
;  APPLICANT: Duke University
;  APPLICANT:  Lin, Haifan
;  TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
;  TITLE OF INVENTION:  PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
;  FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2
;  CURRENT APPLICATION NUMBER: US/09/873,737A
;  CURRENT FILING DATE:  2001-06-04
;  PRIOR APPLICATION NUMBER: PCT/US99/28764
;  PRIOR FILING DATE: 1999-12-03
;  PRIOR APPLICATION NUMBER: 60/110,901
;  PRIOR FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 4064
;   TYPE: DNA
;   ORGANISM: Mus sp.
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (191)..(2776)
;   NAME/KEY: misc_feature
;   LOCATION: (458)
;   OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (836)
;   OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (1337)
;   OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (2636)
;   OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
US-09-873-737A-3

  Query Match           52.8%;  Score 1833.6;  DB 10;  Length 4064;
  Best Local Similarity  81.8%;  Pred. No. 0;
  Matches 2125;  Conservative    0;  Mismatches  471;  Indels    3;  Gaps    1;

Qy        164 ACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCG 223
              | |||||||||  ||||||||||  | ||| | || ||||| || ||||||||||| | |
Db        188 AAAATGACTGGCCGAGCCCGAGCTCGGGCCCGCGGCAGGGCACGAGGTCAGGAGACGGTG 247

Qy        224 CAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCG 283
              ||||  || ||  |  |||| || |||||||||||| || || |  || || ||||| | |
Db        248 CAGCATGTTGGGGCTGCTGCGAGCCAGCAACCTGGGTACATCCCACCGAGACCTCAACAG 307

Qy        284 CCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAA---CAGCAGGAGGA 340
               | ||  |||||||||| ||  ||||||| ||||| ||||| || |    ||  ||||
Db        308 TCCCCCACAGAGGGGGACTTGGTTGGCCGAGGACGACAGAGGGGGATGGTAGTCGGAGCC 367

Qy        341 ACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAG 400
              ||| |||||||||||| ||| ||||| || ||||| ||||||||| | ||  | ||||||
Db        368 ACATCCAAGTCACAAGAACTGCAGATCTCAGCTGGGTTTCAGGAGCTGTCACTGGCAGAG 427

Qy        401 AGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGAC 460
              |||||||| |||||  |||| || |||||  |||||||||| || || |||||||| |||
Db        428 AGAGGAGGGCGTCGCCGAGACTTCCATGACNTTGGTGTGAACACCAGACAGAACCTTGAC 487

Qy        461 CATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCAT 520
              ||||| ||||| ||||| ||||| || || ||||| || || | | | ||||| |||||
Db        488 CATGTCAAAGAGTCAAAGACAGGCTCCTCTGGCATCATTGTGAAGCTGAGCACCAACCAC 547

Qy        521 TTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCA 580
              |||||||||||  || || || |||||||||  | ||||||||  ||||| ||||| || ||
Db        548 TTCCGGCTGACCTCGCGCCCACAGTGGGCCCTGTATCAGTACCACATCGACTACAATCCC 607

Qy        581 CTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGA 640
              |||||||| ||| |||| || || || || || || || || || || ||||| || ||||||
Db        608 CTGATGGAGGCCCGAAGGCTTCGCTCCGCACTGCTCTTCCAGCATGAAGACCTCATTGGA 667

Qy        641 AAGTGTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTT 700
              | |||||||||||| ||||| || ||||| |||||||||| |||||||||| |||||
Db        668 AGGTGTCATGCTTTCGATGGGACAATATTGTTTTTACCTAAGAGACTACAGCACAAGGTC 727

Qy        701 ACTGAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACA 760
              || ||||| || ||| |||| |||||||| ||  | |||||||| || |||||  | ||
Db        728 ACAGAAGTATTCAGTCAGACTCGGAATGGGGAACACGTGAGGATCACCATCACCCTGACC 787

Qy        761 AATGAACTTCCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGG 820
              || || || || || || || || || || ||  |||||||||||| ||  | |||||||||
Db        788 AACGAGCTGCCGCCCACCTCGCCCACCTGCCTGCAGTTCTATAACATCNTCTTCAGGAGG 847

Qy        821 CTTTTGAAAATCATGAATTTGCAACAAATTGGACGAAATTATTATAACCCAAATGACCCA 880
              || |||||||||||||||||||||||||||||||||||||||| |||||||| || |||| |||||
Db        848 CTCTTGAAAATCATGAATTTGCAACAAATTGGACGGAATTATTACAATCCAAGTGACCCG 907

Qy        881 ATTGATATTCCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACTACTTCCATCCTTCAG 940
              ||||||||||||  |||||||||||||| ||||| |||||||| || ||||||||||||
Db        908 ATTGATATTCCAAACCACAGGTTGGTGATCTGGCCCGGCTTCACCACCTCCATCCTTCAG 967

Qy        941 TATGAAAACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAAGTGAGACT 1000
              ||||| |||| |||||||||||||| ||||| ||||| || || || || || ||||||
Db        968 TATGAGAACAACATCATGCTCTGCACAGACGTCAGCCACAAGGTGCTCCGCAGCGAGACT 1027

Qy       1001 GTTTTGGATTTCATGTTCAACTTTTATCATCAGACAGAAGAACATAAATTTCAAGAACAA 1060
              ||  | || |||||||||||  | || || |||||||| || || || || || || |||
Db       1028 GTCCTAGACTTCATGTTCAATCTATACCAGCAGACAGAGGAGCACAAGTTCCAGGAGCAA 1087

Qy       1061 GTTTCCAAAGAANTAATAGGTTTAGTTGTTCTTACCAAGTATAACATTAAGACATACAGA 1120
              || || || ||  | |||||  |  | ||||| |||||||| || |  ||||| ||| |
Db       1088 GTGTCGAAGGAGCTCATAGGCCTCATCGTTCTCACCAAGTACAATAACAAGACCTACCGG 1147

Qy       1121 GTGGATGATATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAAAGCCGACGGCTCT 1180
              |||||||| |||||||||||||||||| |||||||||| ||||| || || |||||
Db       1148 GTGGATGACATTGACTGGGACCAGAATCCAAAGAGCACCTTCAAGAAGGCGGATGGCTCG 1207

Qy       1181 GGGGTCAGCTTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAG 1240
              | |||||||||| | || |||||||||||||||||||||| |||||||| ||| |||||
Db       1208 GAGGTCAGCTTCCTGGAGTACTACAGGAAGCAATACAACCAGGAGATCACGGACCTGAAG 1267

Qy       1241 CAGCCTGTCTTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGG 1300
              ||||| ||  |||| ||||| |||||| | ||| | ||||| || || || ||||| ||
Db       1268 CAGCCGGTGCTGGTGAGCCAACCCAAGCGGAGGAGAGGCCCCGGCGGCACCCTGCCTGGC 1327

Qy       1301 CCTGCCATGCTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAAT 1360
              || || ||| |||| ||||| |||||||||||| ||||| || |||||||||||||| |||
Db       1328 CCAGCTATGNTCATCCCTGAACTCTGCTATCTCACAGGCCTGACTGATAAAATGCGCAAT 1387

Qy       1361 GATTTTAACGTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAG 1420
              ||||| || |||||||| ||| | || || || ||  | || || || |||||  |||||
Db       1388 GATTTCAATGTGATGAAGGACCTGGCAGTGCACACGCGGCTGACCCCTGAGCAGCGGCAG 1447

Qy       1421 CGTGAAGTGGGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTT 1480
```

```
Db       1448 CGGGAGGTGGGCCGCCTCATCGACTACATCCACAAGGATGACAATGTGCAGAGAGAGCTT 1507

Qy       1481 CGAGACTGGGGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAA 1540

Db       1508 CGAGACTGGGGCCTGAGCTTCGACTCAAACTTGCTGTCCTTCTCTGGAAGAATCTTACAA 1567

Qy       1541 ACAGAAAAGATTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGG 1600

Db       1568 TCTGAGAAGATCCACCAGGGCGGAAAGACGTTTGATTACAACCCACAATTTGCAGACTGG 1627

Qy       1601 TCCAAAGAAACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTG 1660

Db       1628 TCCAAAGAAACAAGAGGCGCGCCGCTGATCAGCGTGAAGCCATTGGATAACTGGCTGCTG 1687

Qy       1661 ATCTATACGCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTT 1720

Db       1688 ATCTATACCCGCAGGAATTATGAAGCAGCCAACTCACTGATACAGAACCTGTTCAAAGTG 1747

Qy       1721 ACACCAGCCATGGGCATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACT 1780

Db       1748 ACTCCAGCCATGGGCATCCAGATGAAAAAGGCAATCATGATCGAGGTGGATGACAGAACA 1807

Qy       1781 GAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGT 1840

Db       1808 GAAGCTTATCTGAGAGCCTTGCAGCAGAAGGTGACGTCAGACACTCAGATAGTTGTCTGT 1867

Qy       1841 CTGTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGAT 1900

Db       1868 CTCTTGTCAAGTAATCGGAAGGACAAATATGATGCCATCAAGAAGTACTTGTGTACAGAC 1927

Qy       1901 TGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATG 1960

Db       1928 TGCCCCACCCCAAGTCAGTGTGTGGTGGCCCGGACCCTGGGCAAGCAGCAAACAGTCATG 1987

Qy       1961 GCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGG 2020

Db       1988 GCCATTGCCACCAAGATCGCCCTGCAGATGAACTGCAAGATGGGAGGCGAGCTCTGGCGG 2047

Qy       2021 GTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACA 2080

Db       2048 GTGGACATGGCCCTGAAACTGGCAATGATCGTGGGCATCGACTGTTACCATGACACCACA 2107

Qy       2081 GCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGG 2140

Db       2108 GCTGGGCGGAGGTCCATCGCAGGATTCGTCGCCAGCATCAATGAAGGGATGACCCGCTGG 2167

Qy       2141 TTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGC 2200

Db       2168 TTCTCCCGCTGCGTCTTTCAGGACCGCGGGCAGGAGCTGGTGGATGGTCTCAAGGTGTGC 2227

Qy       2201 CTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATC 2260

Db       2228 TTGCAAGCTGCTCTGAGGGCTTGGAGTGGCTGCAATGAATACATGCCCAGCCGTGTCATC 2287

Qy       2261 GTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCA 2320

Db       2288 GTGTACCGAGACGGTGTGGGGGACGGGCAGCTGAAGACCCTGGTCAATTATGAGGTCCCA 2347

Qy       2321 CAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATT 2380

Db       2348 CAGTTCCTAGATTGCCTCAAGTCAGTCGGGAGAGGTTACAACCCAAGACTGACTGTAATC 2407

Qy       2381 GTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAAT 2440

Db       2408 GTGGTGAAGAAGCGTGTCAATGCCAGGTTTTTTGCTCAGTCTGGGGGAAGACTTCAGAAC 2467

Qy       2441 CCACTTCCTGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTT 2500

Db       2468 CCTCTTCCAGGGACAGTCATCGATGTGGAAGTCACCAGACCAGAGTGGTATGACTTTTTC 2527

Qy       2501 ATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTAT 2560

Db       2528 ATCGTGAGCCAGGCAGTGAGAAGCGGGAGTGTGTCCCCAACACACTACAATGTCATCTAT 2587

Qy       2561 GACAACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATC 2620

Db       2588 GACAGCAGTGGCCTGAAGCCCGACCACATCCAGCGGCTGACATACAAGNTCTGCCACGTG 2647

Qy       2621 TATTACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTG 2680

Db       2648 TACTATAATTGGCCTGGAGTCATCCGAGTCCCTGCACCTTGCCAGTATGCACACAAGCTG 2707

Qy       2681 GCTTTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTT 2740

Db       2708 GCCTTCCTCGTGGGCCAGAGCATCCACAGAGAGCCAAACCTCTCCCTGTCCAACCGCCTC 2767

Qy       2741 TACTACCTCTAACCTGCAG 2759

Db       2768 TACTACCTCTAACCAGCAG 2786
```

RESULT 5
US-09-998-598-2600/c
; Sequence 2600, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 2600
; LENGTH: 582
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 527,531,553
; OTHER INFORMATION: n = A,T,C or G
US-09-998-598-2600

Query Match 15.9%; Score 552.2; DB 11; Length 582;
Best Local Similarity 99.3%; Pred. No. 1.2e-122;
Matches 551; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```
Qy       2915 TTATTTTATATGTAAAAATTAAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTT 2974

Db        582 TTATTTTATATGTAAAAATTAAGATTTTAWATTTTATCTTCTTGTTTCTCAWAGAWATTT 523

Qy       2975 TGTGAGCATTTTTTTTGTTTATTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAAC 3034

Db        522 TGTGAGCATTTTTTTTGTTTATTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAAC 463

Qy       3035 AGACTCTCTGAGAGTATTTGAAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGT 3094

Db        462 AGACTCTCTGAGAGTATTTGAAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGT 403

Qy       3095 GAGCAAGTCCTACTTATAAACCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTT 3154

Db        402 GAGCAAGTCCTACTTATAAACCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTT 343

Qy       3155 AAAGGAGATAAGAGGCGTAAAGTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTC 3214

Db        342 AAAGGAGATAAGAGGCGTAAAGTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTC 283

Qy       3215 ATATCTTAAAGATAAAAGGTACTATTATATAACCTATACACAAGATACAGGAGAAAATAT 3274

Db        282 ATATCTTAAAGATAAAAGGTACTATTATATAACCTATACACAAGATACAGGAGAAAATAT 223
```

```
Qy       3275 GCTTGATTTTTATTTGGCAGGGGGGCTAGGTTGTATGGGAGTAAAAAAAACATTGAAAAT 3334
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        222 GCTTGATTTTTATTTGGCAGGGGGGCTAGGTTGTATGGGAGTAAAAAAAACATTGAAAAT 163

Qy       3335 TTTTAAATTGTCCAAAGAAACATTTTAAGACTCTTTAACAAAAAAGGCCATGAGTAAATC 3394
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        162 TTTTAAATTGTCCAAAGAAACATTTTAAGACTCTTTAACAAAAAAGGCCATGAGTAAATC 103

Qy       3395 TCTATATTAACATCACTATTTATTTTGTTTTGGAACTGGGACATGATTCTATTTGTTATA 3454
             ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db        102 TCTATATTAACATTACTATTTATTTTGTTTTGGAACTGGGACATGATTCTATTTGTTATA 43

Qy       3455 AAATAAAATTGATGT 3469
             ||||||||||||||||
Db         42 AAATAAAATTGATGT 28
```

RESULT 6
US-09-919-580-634/c
; Sequence 634, Application US/09919580
; Patent No. US20020110832A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.552
; CURRENT APPLICATION NUMBER: US/09/919,580
; CURRENT FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 634
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 535, 539, 561
; OTHER INFORMATION: n = A,T,C or G
US-09-919-580-634

Query Match 15.9%; Score 550.4; DB 11; Length 590;
Best Local Similarity 99.3%; Pred. No. 3.3e-122;
Matches 551; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy       2915 TTATTTTATATGTAAAAATTAAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTT 2974
             |||||||||||||||||||||||||||| |||||||||||||||||||| ||| ||||
Db        590 TTATTTTATATGTAAAAATTAAGATTTTANATTTTATCTTCTTGTTTCTCANAGANATTT 531

Qy       2975 TGTGAGCATTTTTTTGTTTATTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAAC 3034
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        530 TGTGAGCATTTTTTTGTTTATTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAAC 471

Qy       3035 AGACTCTCTGAGAGTATTTGAAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGT 3094
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        470 AGACTCTCTGAGAGTATTTGAAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGT 411

Qy       3095 GAGCAAGTCCTACTTATAAACCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTT 3154
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        410 GAGCAAGTCCTACTTATAAACCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTT 351

Qy       3155 AAAGGAGATAAGAGGCGTAAAGTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTC 3214
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        350 AAAGGAGATAAGAGGCGTAAAGTAGGATGCTCACTACAACCATAGGTGGGGTTTCAGCTC 291

Qy       3215 ATATCTTAAAGATAAAAGGTACTATTATATAACCTATACACAAGATACAGGAGAAAATAT 3274
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        290 ATATCTTAAAGATAAAAGGTACTATTATATAACCTATACACAAGATACAGGAGAAAATAT 231

Qy       3275 GCTTGATTTTTATTTGGCAGGGGGGCTAGGTTGTATGGGAGTAAAAAAAACATTGAAAAT 3334
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        230 GCTTGATTTTTATTTGGCAGGGGGGCTAGGTTGTATGGGAGTAAAAAAAACATTGAAAAT 171
```

```
Qy       3335 TTTTAAATTGTCCAAAGAAACATTTTAAGACTCTTTAACAAAAAAGGCCATGAGTAAATC 3394
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        170 TTTTAAATTGTCCAAAGAAACATTTTAAGACTCTTTAACAAAAAAGGCCATGAGTAAATC 111

Qy       3395 TCTATATTAACATCACTATTTATTTTGTTTTGGAACTGGGACATGATTCTATTTGTTATA 3454
             ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db        110 TCTATATTAACATTACTATTTATTTTGTTTTGGAACTGGGACATGATTCTATTTGTTATA 51

Qy       3455 AAATAAAATTGATGT 3469
             ||||||||||||||||
Db         50 AAATAAAATTGATGT 36
```

RESULT 7
US-09-920-300A-34
; Sequence 34, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-920-300A-34

Query Match 13.2%; Score 459.4; DB 11; Length 501;
Best Local Similarity 99.8%; Pred. No. 2.5e-100;
Matches 460; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       2695 CCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAACC 2754
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db          1 CCAGAGTATTCACAGAGAGCCAAATCTGTCACTGGCAAACCGCCTTTACTACCTCTAACC 60

Qy       2755 TGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTT 2814
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTT 120

Qy       2815 TATTTACTTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGA 2874
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TATTTACTTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGA 180

Qy       2875 TCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATATGTAAAAATT 2934
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATATGTAAAAATT 240

Qy       2935 AAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTGTTTA 2994
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTGTTTA 300

Qy       2995 TTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTG 3054
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTG 360

Qy       3055 AAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAAA 3114
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAAA 420

Qy       3115 CCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTTA 3155
             |||||||||||||||||||||||||||||||||||||||||
Db        421 CCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTTA 461
```

```
RESULT 8
US-10-033-528-34
; Sequence 34, Application US/10033528
; Publication No. US20020131971A1
; GENERAL INFORMATION:
;  APPLICANT: King, Gordon E.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Secrist, Heather
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.547C1
;  CURRENT APPLICATION NUMBER: US/10/033,528
;  CURRENT FILING DATE:  2001-12-26
;  NUMBER OF SEQ ID NOS: 1896
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34
;   LENGTH: 501
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-033-528-34

  Query Match             13.2%;  Score 459.4;  DB 15;  Length 501;
  Best Local Similarity   99.8%;  Pred. No. 2.5e-100;
  Matches  460;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2695 CCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAACC 2754
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db          1 CCAGAGTATTCACAGAGAGCCAAATCTGTCACTGGCAAACCGCCTTTACTACCTCTAACC 60

Qy       2755 TGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTT 2814
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTT 120

Qy       2815 TATTTACTTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGA 2874
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TATTTACTTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGA 180

Qy       2875 TCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATATGTAAAAATT 2934
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATATGTAAAAATT 240

Qy       2935 AAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTGTTTA 2994
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTGTTTA 300

Qy       2995 TTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTG 3054
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTG 360

Qy       3055 AAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAAA 3114
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAAA 420

Qy       3115 CCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTTA 3155
             |||||||||||||||||||||||||||||||||||||||||
Db        421 CCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTTA 461


RESULT 9
US-09-777-564-1706/c
; Sequence 1706, Application US/09777564
; Patent No. US20020022591A1
; GENERAL INFORMATION:
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Mannion, Jane
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE:  210121.493
;  CURRENT APPLICATION NUMBER: US/09/777,564
;  CURRENT FILING DATE:  2001-02-05
;  NUMBER OF SEQ ID NOS:  1730
;  SOFTWARE:  FastSEQ for Window Version 4.0
; SEQ ID NO 1706
;   LENGTH: 460
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-777-564-1706

  Query Match             13.2%;  Score 458.4;  DB 10;  Length 460;
  Best Local Similarity   99.8%;  Pred. No. 4.1e-100;
  Matches  459;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2695 CCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAACC 2754
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 CCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAACC 401

Qy       2755 TGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTT 2814
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 TGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTT 341

Qy       2815 TATTTACTTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGA 2874
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        340 TATTTACTTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGA 281

Qy       2875 TCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATATGTAAAAATT 2934
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        280 TCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATACGTAAAAATT 221

Qy       2935 AAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTGTTTA 2994
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 AAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTGTTTA 161

Qy       2995 TTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTG 3054
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        160 TTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTG 101

Qy       3055 AAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAAA 3114
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        100 AAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAAA 41

Qy       3115 CCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTT 3154
             |||||||||||||||||||||||||||||||||||||||
Db         40 CCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTT 1


RESULT 10
US-10-015-219-1706/c
; Sequence 1706, Application US/10015219
; Publication No. US20030008299A1
; GENERAL INFORMATION:
;  APPLICANT: Algate, Paul A.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.493C1
;  CURRENT APPLICATION NUMBER: US/10/015,219
;  CURRENT FILING DATE:  2002-03-02
;  NUMBER OF SEQ ID NOS: 1739
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1706
;   LENGTH: 460
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-015-219-1706

  Query Match             13.2%;  Score 458.4;  DB 15;  Length 460;
  Best Local Similarity   99.8%;  Pred. No. 4.1e-100;
  Matches  459;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2695 CCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAACC 2754
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 CCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAACC 401

Qy       2755 TGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTT 2814
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 TGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTTTTTTAAGCTTT 341
```

```
Qy       2815 TATTTACTTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGA 2874
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        340 TATTTACTTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGGGGATTAGGAGA 281

Qy       2875 TCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATATGTAAAAATT 2934
             ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        280 TCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTATACGTAAAAATT 221

Qy       2935 AAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTGTTTA 2994
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 AAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCATTTTTTTGTTTA 161

Qy       2995 TTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTG 3054
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        160 TTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTCTGAGAGTATTTG 101

Qy       3055 AAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAAA 3114
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        100 AAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAGTCCTACTTATAAA 41

Qy       3115 CCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTT 3154
             |||||||||||||||||||||||||||||||||||||||||
Db          40 CCTATATTAACTTTATTTTTGAGATACCTGTTTTGAATTT 1
```

RESULT 11
US-10-029-495-8
; Sequence 8, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
; APPLICANT: Drmanac, Radoje T.
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Dickson, Mark C.
; APPLICANT: Jones, Lee William
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Xue, Aidong
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi,Vinod
; TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
; FILE REFERENCE: 28110/35905A
; CURRENT APPLICATION NUMBER: US/10/029,495
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 09/378,667
; PRIOR FILING DATE: 1999-08-20
; PRIOR APPLICATION NUMBER: 09/687,527
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 2009
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (73)..(1569)
US-10-029-495-8

Query Match 13.2%; Score 457.4; DB 15; Length 2009;
Best Local Similarity 57.1%; Pred. No. 1.7e-99;
Matches 893; Conservative 0; Mismatches 662; Indels 9; Gaps 3;

```
Qy       1189 CTTCTTAGAATACTACAGGAAGCAATACAACCAAGAGATCACCGACTTGAAGCAGCCTGT 1248
              ||   | || ||||||| | |||| ||  |        |  | ||| |||| |||||  |
Db         15 CTAGGTGGATTACTACAAGCAGCAGTATGATATTACTGTATCGGACCTGAATCAGCCCAT 74

Qy       1249 CTTGGTCAGCCAGCCCAAGAGAAGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCAT 1308
              | || || | |   |||   | |  | |   ||      |    |||   ||  |||
Db         75 GCTTGTTAGTCTGTTAAAG---AAGAAGAGAAATGACAACAGTGAGCCTCAGCTCGCCCA 131

Qy       1309 GCTCATTCCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAA 1368
```

```
Db        132 CCTGATACCTGAGCTCTGCTTTCTAACAGGGCTGACTGACCAGGCAACATCTGATTTCCA 191

Qy       1369 CGTGATGAAAGACTTAGCCGTTCATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGT 1428
              |||||||  |   | || |   | ||| | || | |||          ||||||     |
Db        192 GCTGATGAAGGCTGTGGCTGAAAAGACACGTCTCAGTCCTTCAGGCCGGCAGCAGCGCCT 251

Qy       1429 GGGACGACTCATTGATTACATTCATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTG 1488
              ||   | ||  | ||  |||| || |  ||    |||| ||      || ||      |||
Db        252 GGCCAGGCTTGTGGACAACATCCAGAGGAATACCAATGCTCGCTTTGAACTAGAGACCTG 311

Qy       1489 GGGTTTGAGCTTTGATTCCAACTTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAA 1548
              |||  ||  ||||    | |  ||  ||  | | ||  | ||| |||   |||||||
Db        312 GGGACTGCATTTTGGAAGCCAGATA---TCTCTGACTGGCCGGATTGTGCCTTCAGAAAA 368

Qy       1549 GATTCACCAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCCAAAGA 1608
              ||           |   | | ||    | |  |        ||| || ||||||||  ||
Db        369 AATATTAATGCAAGACCACATAT---GTCAACCTGTGTCTGCTGCTGACTGGTCCAAGGA 425

Qy       1609 AACAAGAGGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATAC 1668
              |   ||  |       | | | ||   || | |  ||| |||| || |  | | ||
Db        426 TATTCGAACTTGCAAGATTTTAAATGCACAGTCTTTGAATACCTGGTTGATTTTATGTAG 485

Qy       1669 GCGAAGAAATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACCAGC 1728
                 |||| | |  | |  ||| |    ||  |  |     |    | |||| ||    |
Db        486 CGACAGAACTGAATATGTTGCCGAGAGCTTTCTGAACTGCTTGAGAAGAGTTGCAGGTTC 545

Qy       1729 CATGGGCATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACTGAAGCCTA 1788
              ||||||  |  |  ||    |  | | ||| ||  ||||  | || |  | |  || |
Db        546 CATGGGATTTAATGTGGACTACCCCAAAATCATAAAAGTACAAGAAAATCCAGCTGCATT 605

Qy       1789 CTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCTGTTGTC 1848
              | ||||  ||||||||  | ||     | ||    ||| | ||  | ||  |  || |
Db        606 TGTTAGAGCTATACAGCAATATGTTGATCCTGATGTTCAGCTGGTAATGTGCATTCTGCC 665

Qy       1849 AAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTGCCCTAC 1908
              |||| ||||  | | || ||| | ||||||||||  || |  |||| |||| |||||
Db        666 TTCTAATCAGAAGACCTATTATGATTCCATTAAAAAATATTTGAGCTCAGACTGCCCAGT 725

Qy       1909 CCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGC 1968
              |||||| || |||||| | || || |||||    ||||||   |   | |||   || ||
Db        726 CCCAAGCCAATGTGTGCTTGCTCGGACCTTGAATAAACAGGGCATGATGATGAGTATCGC 785

Qy       1969 TACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACAT 2028
              || ||||| ||  | |||||||  |||||| | |||||  |||| |||   ||||| ||
Db        786 CACCAAGATCGCTATGCAGATGACTTGCAAGCTCGGAGGCGAGCTGTGGGCTGTGGAAAT 845

Qy       2029 CCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCG 2088
              ||  | |||  | ||||| | || || || |||   |   | ||                |
Db        846 ACCTTTAAAGTCCCTGATGGTGGTCGGTATTGATGTCTGTAAAGATGCACTCAGCAAGGA 905

Qy       2089 GAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACG 2148
                 |  | |  |||| || |||||  | ||     | || ||| | ||||| || ||
Db        906 CGTGATGGTTGTTGGATGCGTGGCCAGTGTTAACCCCAGAATCACCAGGTGGTTTTCCCG 965

Qy       2149 CTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGC 2208
              ||| ||  |||||   | |      ||  | | |||| |  | ||||| | | ||   |
Db        966 CTGTATCCTTCAGAGAACAATGACTGATGTTGCAGATTGCTTGAAAGTTTTCATGACTGG 1025

Qy       2209 GGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATCGTGTACCG 2268
              || || |     ||| | | | |||| |  |  ||||    ||||| || ||||||||
Db        1026 AGCACTCAACAAATGGTACAAGTACAATCATGATTTGCCAGCACGGATAATTGTGTACCG 1085

Qy       2269 CGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCACAGTTTTT 2328
              | ||| ||||| || || ||||||||||||||  |  | || ||||| |||||| | |
Db        1086 TGCTGGTGTAGGGGATGGTCAGCTGAAAACACTTATTGAATATGAAGTCCCACAGCTGCT 1145

Qy       2329 GGATTGTCTAAAATCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATTGTGGTGAA 2388
              |    || |  |       |  | |  || | | ||| |  |||| |||||||| |
Db        1146 GAGCAGTGTGGCAGAATCCAGCTCAAATACCAGCTCAAGACTGTCGGTGATTGTGGTCAG 1205

Qy       2389 GAAAAGAGTGAACACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTCC 2448
              ||| |     |  |  |||| ||| |  |        | |   | ||||| || |  |
```

```
Db       1206 GAAGAAGTGCATGCCACGATTCTTTACCGAAATGAACCGCACTGTACAGAACCCCCCACT 1265

Qy       2449 TGGAACAGTTATTGATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTTATCGTGAG 2508

Db       1266 TGGCACTGTTGTGGATTCAGAAGCAACACGTAACGAATGGTATGACTTTTATCTGATCAG 1325

Qy       2509 CCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTATGACAACAG 2568

Db       1326 CCAGGTGGCCTGCCGGGGAACTGTTAGTCCTACCTACTATAATGTCATCTATGATGACAA 1385

Qy       2569 CGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATCTATTACAA 2628

Db       1386 CGGCTTGAAGCCCGACCATATGCAGAGACTTACATTCAAATTGTGCCACCTGTACTACAA 1445

Qy       2629 CTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCT 2688

Db       1446 CTGGCCGGGCATAGTCAGTGTCCCAGCACCATGTCAGTATGCTCACAAGCTGACCTTTCT 1505

Qy       2689 TGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCT 2748

Db       1506 GGGGGCACAAAGCATTCATAAAGAACCCAGTCTGGAATTAGCCAACCATCTCTTCTACCT 1565

Qy       2749 CTAA 2752

Db       1566 GTGA 1569


RESULT 12
US-10-102-806-267
; Sequence 267, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA103P1C1
;  CURRENT APPLICATION NUMBER: US/10/102,806
;  CURRENT FILING DATE:  2002-03-22
;  PRIOR APPLICATION NUMBER: 09/925,298
;  PRIOR FILING DATE: 2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05881
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 846
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 267
;   LENGTH: 423
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-102-806-267

  Query Match             11.4%;  Score 395.4;  DB 15;  Length 423;
  Best Local Similarity   98.8%;  Pred. No. 5.7e-85;
  Matches  404;  Conservative    4;  Mismatches    0;  Indels    1;  Gaps    1;

Qy       1898 GATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTC 1957

Db         14 GATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTC 73

Qy       1958 ATGGCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGG 2017

Db         74 ATGGCCATTGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGGRGAGCTCTGG 133

Qy       2018 AGGGTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATG 2077

Db        134 AGGGTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTKACCATGACATG 193

Qy       2078 ACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGC 2137

Db        194 ACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGC 253

Qy       2138 TGGTTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTC 2197

Db        254 TGGTTCTCACGCTGCATATTTCAGGATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTC 313

Qy       2198 TGCCTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATC 2257

Db        314 TGCCTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCAGCCGGATC 373

Qy       2258 ATCGTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGA 2306

Db        374 ATCGTGTACCGSG-TGGCGTAGGAGACGGCCAGYTGAAAACACTGGTGA 421


RESULT 13
US-09-801-574-15
; Sequence 15, Application US/09801574
; Patent No. US20020081592A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Peijing Jeremy
;  APPLICANT:  Page, David C.
;  TITLE OF INVENTION: Reproduction-Specific Genes
;  FILE REFERENCE: 0399.2007-002
;  CURRENT APPLICATION NUMBER: US/09/801,574
;  CURRENT FILING DATE:  2001-03-07
;  PRIOR APPLICATION NUMBER: 60/187,518
;  PRIOR FILING DATE: 2000-03-07
;  PRIOR APPLICATION NUMBER: 60/261,557
;  PRIOR FILING DATE: 2001-01-12
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
;   LENGTH: 3649
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-801-574-15

  Query Match             10.9%;  Score 377;  DB 10;  Length 3649;
  Best Local Similarity   52.5%;  Pred. No. 5.7e-80;
  Matches  974;  Conservative    0;  Mismatches  862;  Indels   21;  Gaps    6;

Qy        914 CCTGGCTTCACTACTTCCATCCTTCAGTATGAAAACAGCATCATGCTCTGCACTGACGTT 973

Db          1 CCTGGCTATGCGGCTAGTATCCGGAGGACAGACGGGGGTCTCTTCCTGCTCGCTGATGTC 60

Qy        974 AGCCATAAAGTCCTTCGAAGTGAGACTGTTTTGGATTTCATGTTCAACTTTTATCATCAG 1033

Db         61 TCTCATAAGGTCATTCGGAACGACTCTGTGCTGGATGTCATGCATGCTATCTACCAGCAG 120

Qy       1034 ACAGAAGAACATAAATTTCAAGAACAAGTTTCCAAAGAANTAATAGGTTTAGTTGTTCTT 1093

Db        121 AACAAGGAGCAC---TTCCAGGACGAGTGCAGCAAGCTTCTGGTTGGCAGCATTGTCATC 177

Qy       1094 ACCAAGTATAACATTAAGACATACAGAGTGGATGATATTGACTGGGACCAGAATCCCAAG 1153

Db        178 ACGCGCTACAACAATCGTACCTACCGAATCGATGATGTGGACTGGAACAAGACCCCTAAA 237

Qy       1154 AGCACCTTTAAGAAAGCCGACGGCTCTGGGGTCAGCTTCTTAGAATACTACAGGAAGCAA 1213

Db        238 GACAGCTTTGTCATGTCGGACGGGAAGGAAATCACATTCCTGGAATACTACAGCAAAAAC 297

Qy       1214 TACAACCAAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGCCAGCCCA---AGAGA 1270

Db        298 TATGGGATCACAGTCAAGGAAGATGACCAGCCGCTGCTGATCCACCGGCCCAGTGAGAGA 357

Qy       1271 AGGCGGGGCCCTGGGGGGACACTGCCAGGGCCTGCCATGCTCATTCCTGAGCTCTGCTAT 1330

Db        358 CAGAATAACCATGGCATGTTGCTGAAGGGCGAGATCCTGCTGCTGCCCGAGCTCTCCTTC 417

Qy       1331 CTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAACGTGATGAAAGACTTAGCCGTT 1390

Db        418 ATGACGGGGATCCCTGAGAAGATGAAGAAGGACTTCAGGGCCATGAAGGACTTGACTCAG 477

Qy       1391 CATACAAGACTAACTCCAGAGCAAAGGCAGCGTGAAGTGGGACGACTCATTGATTACATT 1450

Db        478 CAGATTAACCTGAGCCCCAAGCAGCACCACGGTGCTTTGGAATGCCTGCTGCAGAGAATT 537
```

```
Qy       1451 CATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGGGTTTGAGCTTTGATTCCAAC 1510
Db        538 TCACAAAACGAGACAGCCAGCAATGAGCTGACCCGCTGGGGGCTCAGTCTGCATAAAGAT 597

Qy       1511 TTACTGTCCTTCTCAGGAAGAATTTTGCAAACAGAAAAGATTCACCAAGGTGGAAAAACA 1570
Db        598 GTCCACAAGATTGAAGGTCGGCTTCTGCCAATGGAGAGGATCAACTTAAGGAACACTTCA 657

Qy       1571 TTTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAACAAGAGGTGCACCATTAATT 1630
Db        658 TTTGT---CACATCGGAGGGCCTGAACTGGGTTAAGGAAGTGACCAGAGATGCTTCCATT 714

Qy       1631 AGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACGCGAAGAAATTATGAAGCAGCC 1690
Db        715 CTAACTATTCCCATGCATTTCTGGGCACTCTTTTATCCAAAGAGAGCAATGGACCAAGCC 774

Qy       1691 AATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGCATGCAAATGAGAAAA 1750
Db        775 AGAGAACTGGTTAACATGTTGGAAAAGATTGCCGGGCCCATTGGCATGCGCACAAGCCCC 834

Qy       1751 GCAATAATGATTGAAGT---GGATGACAGAACTGAAGCCTACTTAAGAGTCTTACAGCAA 1807
Db        835 CCAGCCTGGGTTGAGCTGAAGGATGACCGAATAGAGACCTATATCAGGACCATTCAGTCC 894

Qy       1808 AAGGTCACAGCAGAC------ACCCAGATAGTTGTCTGTCTGTTGTCAAGTAATCGGAAG 1861
Db        895 TTACTGGGAGTTGAGGGGAAGATACAAATGGTCGTTTGCATCATCATGGGCACACGTGAT 954

Qy       1862 GACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAGTGT 1921
Db        955 GATCTCTATGGAGCCATCAAGAAGCTGTGCTGCGTGCAGTCCCCAGTGCCCTCACAGGTC 1014

Qy       1922 GTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAAAGATTGCC 1981
Db        1015 ATCAATGTCCGAACCATTGGTCAGCCCACCAGGCTTCGGAGCGTGGCTCAGAAAATTTTA 1074

Qy       1982 CTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCTC 2041
Db        1075 CTTCAGATGAACTGTAAACTGGGTGGTGAGCTCTGGGGAGTGGATATTCCGCTGAAACAA 1134

Qy       2042 GTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGCA 2101
Db        1135 CTAATGGTGATTGGAATGGATGTGTACCATGACCCCAGCAGAGGCATGCGCTCTGTGGTC 1194

Qy       2102 GGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCATATTTCAG 2161
Db        1195 GGCTTCGTGGCCAGCATAAATCTCACACTCACCAAATGGTACTCGAGGGTGGTGTTCCAG 1254

Qy       2162 GATAGAGGACAGGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGGCTCTGAGGGCT 2221
Db        1255 ATGCCACATCAGGAGATTGTGGACAGCCTGAAGCTCTGCCTGGTGGGTTCCTTGAAAAAG 1314

Qy       2222 TGGAATAGCTGCAATGAGTACATGCCCAGCCGGATCATCGTGTACCGCGATGGCGTAGGA 2281
Db        1315 TATTATGAGGTGAACCATTGTCTCCCAGAGAAAATTGTGGTGTACCGAGATGGAGTGTCT 1374

Qy       2282 GACGGCCAGCTGAAAACACTGGTGAACTACGAAGTGCCACAGTTTTTGGATTGTCTAAAA 2341
Db        1375 GATGGCCAGCTAAAGACAGTTGCCAACTACGAGATCCCTCAGCTGCAGAAGTGTTTTGAA 1434

Qy       2342 TCCATTGGTAGAGGTTACAACCCTAGANTAACGGTAATTGTGGTGAAGAAAAGAGTGAAC 2401
Db        1435 GCCTTTGATA---ACTACCACCCCAAGATGGTGGTGTTTGTAGTTCAGAAGAAAATCAGC 1491

Qy       2402 ACCAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAATCCACTTCCTGGAACAGTTATT 2461
Db        1492 ACCAATCTGTACCTTGCTGCTCCTGATCACTTCGTAACCCCCTCCCCCGGGACTGTGGTT 1551

Qy       2462 GATGTAGAGGTTACCAGACCAGAATGGTATGACTTTTTTATCGTGAGCCAGGCTGTGAGA 2521
Db        1552 GATCATACCATAACCAGCTGTGAGTGGGTGGATTTCTACCTTCTTGCCCATCATGTGCGA 1611

Qy       2522 AGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTATGACAACAGCGGCCTGAAGCCA 2581
Db        1612 CAGGGCTGTGGCATACCTACACACTACATCTGTGTTCTGAACACTGCAAATCTGAGCCCT 1671

Qy       2582 GACCACATACAGCGCTTGACCTACAAGCTGTGCCACATCTATTACAACTGGCCAGGTGTC 2641
Db        1672 GATCACATGCAGAGGTTGACTTTCAAACTATGCCACATGTACTGGAATTGGCCTGGTACC 1731

Qy       2642 ATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCTTTTCTTGTTGGCCAGAGT 2701
Db        1732 ATCCGAGTTCCAGCTCCTTGCAAGTATGCCCACAAGCTAGCTTTCCTGTCCGGACAGATT 1791

Qy       2702 ATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTACTACCTCTAACCTGCA 2758
Db        1792 TTGCATCATGAGCCAGCCATCCAGCTGTGTGGGAACCTGTTCTTCCTGTAACTGGGA 1848
```

RESULT 14
US-09-873-737A-1
; Sequence 1, Application US/09873737A
; Patent No. US20020076797A1
; GENERAL INFORMATION:
; APPLICANT: Duke University
; APPLICANT: Lin, Haifan
; TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
; TITLE OF INVENTION: PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
; FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2
; CURRENT APPLICATION NUMBER: US/09/873,737A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/US99/28764
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/110,901
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3047
; TYPE: DNA
; ORGANISM: Drosophila sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (84)..(2612)
; NAME/KEY: misc_feature
; LOCATION: (120)
; OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (399)
; OTHER INFORMATION: n=a or t, Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (2436)
; OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
US-09-873-737A-1

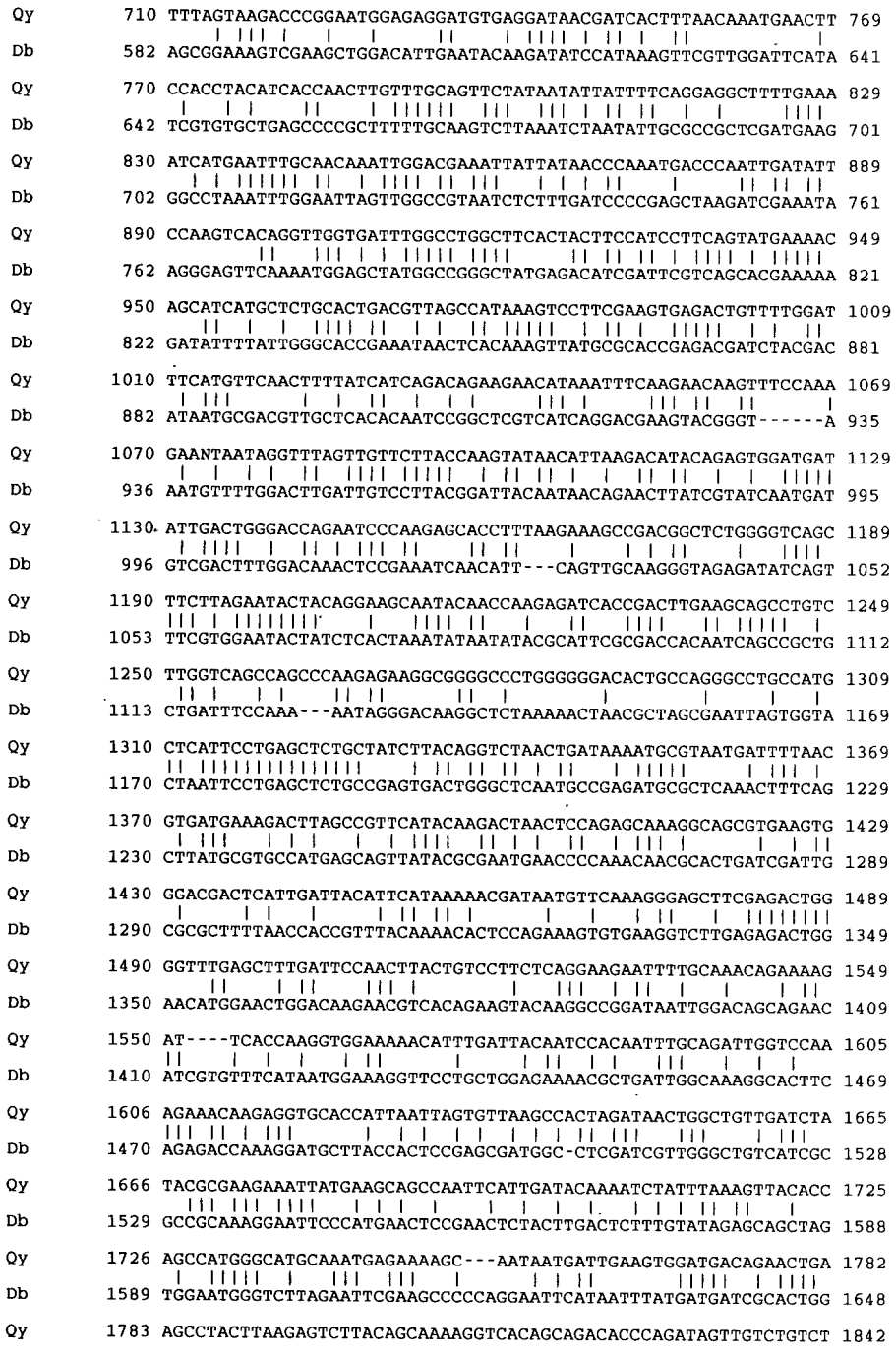
Query Match 7.3%; Score 253; DB 10; Length 3047;
Best Local Similarity 47.1%; Pred. No. 3.7e-50;
Matches 1080; Conservative 0; Mismatches 1184; Indels 29; Gaps 9;

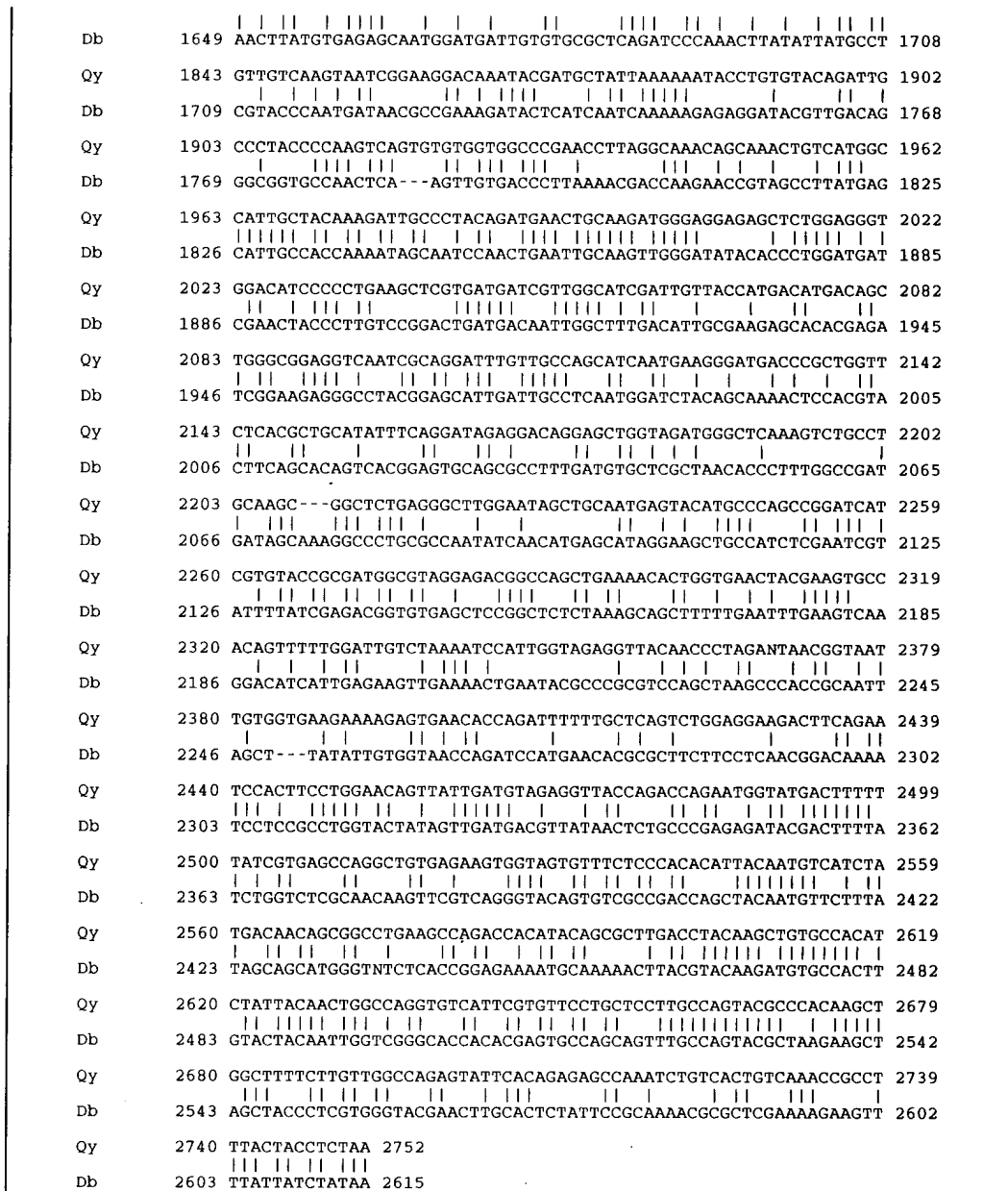
```
Qy        470 GAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTG 529
Db        342 GTATCCAAGAAGGGAACCGATGGCGTCCCGGTCATGCTGCAGACGAACTTTTTTCGANTA 401

Qy        530 ACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGGAA 589
Db        402 AAAACCAAGCCGGAATGGCGGATCGTTCATTATCACGTGGAGTTTGTGCCGACCATCGAG 461

Qy        590 GCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCAT 649
Db        462 AATCCTCGTGTCCGTATGGGAGTTTTGTCCAATCATGCTAACCTTCTGGGATCAGGCTAT 521

Qy        650 GCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTT 709
Db        522 CTATTCGACGGACTGCAACTGTTCACCACCAGGAAATTCGAGCAGGAAATCACGGTGCTC 581
```





RESULT 15
US-10-029-495-1
; Sequence 1, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:

```
;  APPLICANT: Drmanac, Radoje T.
;  APPLICANT:  Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Dickson, Mark C.
;  APPLICANT:  Jones, Lee William
;  APPLICANT:  Ballinger, Dennis G.
;  APPLICANT:  Xue, Aidong
;  APPLICANT:  Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi,Vinod
;  TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
;  FILE REFERENCE: 28110/35905A
;  CURRENT APPLICATION NUMBER: US/10/029,495
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: 09/378,667
;  PRIOR FILING DATE: 1999-08-20
;  PRIOR APPLICATION NUMBER: 09/687,527
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 560
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2)..(559)
;   OTHER INFORMATION: hiwi-Hy/CG 389
US-10-029-495-1

  Query Match             5.5%;  Score 192;  DB 15;  Length 560;
  Best Local Similarity   59.1%;  Pred. No. 6.5e-36;
  Matches  321;  Conservative    3;  Mismatches  219;  Indels    0;  Gaps    0;

Qy       1873 TGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCCG 1932
             | | ||||||||||||  || |  :|:| |||||   |||||| || ||||||:| || ||
Db          1 TTCCATTAAAAAATATTTGAGCTYAKACTGCCCAGTCCCAAGCCAATGTGTGSTTGCTCG 60

Qy       1933 AACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAAAGATTGCCCTACAGATGAA 1992
              |||||    ||||||   |   | |||   || || || ||||| ||  | |||||||
Db         61 GACCTTGAATAAACAGGGCATGATGATGAGTATCGCCACCAAGATCGCTATGCAGATGAC 120

Qy       1993 CTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCTCGTGATGATCGT 2052
              |||||| | ||||| ||||| |||   ||||| || ||  | |||  | ||||| | ||
Db        121 TTGCAAGCTCGGAGGCGAGCTGTGGGCTGTGGAAATACCTTTAAAGTCCCTGATGGTGGT 180

Qy       2053 TGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTGC 2112
              || || |||   |   | ||            |     |    | |  ||||  || ||
Db        181 CGGTATTGATGTCTGTAAAGATGCACTCAGCAAGGACGTGATGGTTGTTGGATGCGTGGC 240

Qy       2113 CAGCATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCATATTTCAGGATAGAGGACA 2172
              |||  | ||     | || ||| | ||||| || ||||| ||  |||||   | |
Db        241 CAGTGTTAACCCCAGAATCACCAGGTGGTTTTCCCGCTGTATCCTTCAGAGAACAATGAC 300

Qy       2173 GGAGCTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGGCTCTGAGGGCTTGGAATAGCTG 2232
              ||  | | |||| |  | ||||| | | ||   |  || || |    ||| | |  |
Db        301 TGATGTTGCAGATTGCTTGAAAGTTTTCATGACTGGAGCACTCAACAAATGGTACAAGTA 360

Qy       2233 CAATGAGTACATGCCCAGCCGGATCATCGTGTACCGCGATGGCGTAGGAGACGGCCAGCT 2292
              |||| |  |  ||||    ||||| || |||||||| | ||| ||||| || || |||||
Db        361 CAATCATGATTTGCCAGCACGGATAATTGTGTACCGTGCTGGTGTAGGGGATGGTCAGCT 420

Qy       2293 GAAAACACTGGTGAACTACGAAGTGCCACAGTTTTTGGATTGTCTAAAATCCATTGGTAG 2352
              |||||||||  |  | || ||||| |||||| |  ||    || |   |       |
Db        421 GAAAACACTTATTGAATATGAAGTCCCACAGCTGCTGAGCAGTGTGGCAGAATCCAGCTC 480

Qy       2353 AGGTTACAACCCTAGANTAACGGTAATTGTGGTGAAGAAAAGAGTGAACACCAGATTTTT 2412
              |  |  || | | ||| |  |||| |||||||| | ||| |     |  |  |||| ||
Db        481 AAATACCAGCTCAAGACTGTCGGTGATTGTGGTCAGGAAGAAGTGCATGCCACGATTCTT 540

Qy       2413 TGC 2415
              | |
Db        541 TAC 543
```

Search completed: July 22, 2003, 06:59:28
Job time : 673.254 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2003, 13:46:42 ; Search time 4559.56 Seconds
(without alignments)
12332.482 Million cell updates/sec

Title: US-09-873-737A-5
Perfect score: 3472
Sequence: 1 gggcgaattgggcccgacgt.....taaaataaaattgatgtccc 3472

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	736.8	21.2	818	12	BG773137	BG773137 602721862
2	730.2	21.0	779	13	BI560710	BI560710 603254744
3	689.2	19.9	1092	13	BI463482	BI463482 603203560
4	672.2	19.4	882	13	BI459464	BI459464 603200456
5	637.2	18.4	715	12	BG721550	BG721550 602695155
6	633	18.2	666	13	BI463897	BI463897 603202635
7	626.2	18.0	731	13	BI458664	BI458664 603199792
8	621.4	17.9	640	9	AL705301	AL705301 DKFZp686G
9	621	17.9	622	9	AL705072	AL705072 DKFZp686E
10	580.6	16.7	718	14	BQ435882	BQ435882 AGENCOURT
11	574.6	16.5	874	12	BF666315	BF666315 602119233
12	560.6	16.1	581	12	BG718168	BG718168 602696127
13	523.2	15.1	539	13	BI463615	BI463615 603206270
14	521	15.0	744	13	BI459142	BI459142 603199321
15	515.2	14.8	577	12	BF080918	BF080918 233684 MA
16	513.4	14.8	527	14	BM751055	BM751055 K-EST0026
17	502.4	14.5	1495	12	BF664941	BF664941 602119191
18	492.8	14.2	605	9	AL704986	AL704986 DKFZp686O
19	492.6	14.2	705	12	BF246708	BF246708 601855507
20	471.2	13.6	533	12	BF080923	BF080923 233690 MA
21	468.2	13.5	830	12	BF699368	BF699368 602125943
c 22	452.4	13.0	463	9	AA639672	AA639672 nq83h10.s
c 23	450.4	13.0	482	10	AW189518	AW189518 xl07g10.x
24	443.4	12.8	474	9	AL705611	AL705611 DKFZp686L
25	442.2	12.7	806	12	BG724366	BG724366 602692837
26	440.4	12.7	462	9	AA969938	AA969938 op60a11.s
c 27	434.4	12.5	446	9	AI125224	AI125224 qd87e12.x
c 28	415.8	12.0	454	9	AL044133	AL044133 DKFZp434O
29	414.2	11.9	565	12	BG101107	BG101107 uy19e03.y
c 30	412	11.9	413	9	AL041473	AL041473 DKFZp434E
c 31	403.4	11.6	417	9	AA969660	AA969660 op39a11.s
32	387.8	11.2	420	14	BQ374910	BQ374910 MR4-TN010
33	387	11.1	477	13	BM539361	BM539361 hb08e02.g
34	386	11.1	587	9	AA430311	AA430311 zw68h01.r
35	373.4	10.8	935	12	BF247005	BF247005 601855750
36	345.2	9.9	485	12	BG718140	BG718140 602698887
37	339	9.8	971	10	BB614499	BB614499 BB614499
c 38	328.2	9.5	349	9	AA904973	AA904973 oj75e06.s
39	326.8	9.4	872	13	BI560401	BI560401 603253963
40	287.2	8.3	609	14	BQ481304	BQ481304 faa74e09.
41	274.8	7.9	559	14	BM728018	BM728018 UI-E-EJ0-
c 42	274.4	7.9	300	9	AI538315	AI538315 tp56a12.x
c 43	265.4	7.6	286	9	AI698821	AI698821 wd31d06.x
c 44	263.8	7.6	512	14	BM680592	BM680592 UI-E-EJ0-
45	259.2	7.5	617	13	BJ040023	BJ040023 BJ040023

ALIGNMENTS

RESULT 1
BG773137
LOCUS BG773137 818 bp mRNA linear EST 15-MAY-2001
DEFINITION 602721862F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4838918 5', mRNA sequence.
ACCESSION BG773137
VERSION BG773137.1 GI:14083790
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 818)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10773 row: k column: 15
High quality sequence stop: 789.
FEATURES Location/Qualifiers

```
    source          1. .818
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:4838918"
                    /clone_lib="NIH_MGC_97"
                    /lab_host="DH10B"
                    /note="Organ: testis; Vector: pBluescriptR (modified
                    pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                    ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                    size-selected for  average insert size 2.2 kb and
                    normalized to ROT 5. This is a primary library enriched
                    for full-length clones and constructed using the
                    Cap-trapper method (Carninci, in preparation).  Library
                    constructed by M. Brownstein (NIMH/NHGRI, National
                    Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      245 a    174 c    198 g    201 t
ORIGIN

  Query Match             21.2%;  Score 736.8;  DB 12;  Length 818;
  Best Local Similarity   98.2%;  Pred. No. 1.6e-146;
  Matches  798;  Conservative    0;  Mismatches    8;  Indels    7;  Gaps    5;

Qy        133 AGGGCGAGGGCAGCGGTCCAAGAAATAGAAAACAATGACTGGGAGAGCCCGAGCCAGAGC 192
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          6 AGGGTGAGGGCAGCGGTCCAAGAAATAGAAAACAATGACTGGGAGAGCCCGAGCCAGAGC 65

Qy        193 CAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGCTGGTGGGCTCCACTGCCAGTCAGCA 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 CAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGCTGGTGGGCTCCACTGCCAGTCAGCA 125

Qy        253 ACCTGGTTATATTCAGCCTAGGCCTCAGCCGCCACCAGCAGAGGGGGGAATTATTTGGCCG 312
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        126 ACCTGGTTATATTCAGCCTAGGCCTCAGCCGCCACCAGCAGAGGGGGGATTATTTGGCCG 185

Qy        313 TGGACGGCAGAGAGGAACAGCAGGAGGAACAGCCAAGTCACAAGGACTCCAGATATCTGC 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 TGGACGGCAGAGAGGAACAGCAGGAGGAACAGCCAAGTCACAAGGACTCCAGATATCTGC 245

Qy        373 TGGATTTCAGGAGTTATCGNTAGCAGAGAGAGGAGGTCGTCGTAGAGATTTTCATGATCT 432
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        246 TGGATTTCAGGAGTTATCGTTAGCAGAGAGAGGAGGTCGTCGTAGAGATTTTCATGATCT 305

Qy        433 TGGTGTGAATACAAGGCAGAACCTAGACCATGTTAAAGAATCAAAAACAGGTTCTTCAGG 492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 TGGTGTGAATACAAGGCAGAACCTAGACCATGTTAAAGAATCAAAAACAGGTTCTTCAGG 365

Qy        493 CATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGACATCCCGTCCCCAGTGGGCCTT 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 CATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGACATCCCGTCCCCAGTGGGCCTT 425

Qy        553 ATATCAGTATCACATTGACTATAACCCACTGATGGAAGCCAGAAGACTCCGTTCAGCTCT 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 ATATCAGTATCACATTGACTATAACCCACTGATGGAAGCCAGAAGACTCCGTTCAGCTCT 485

Qy        613 TCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCATGCTTTTGATGGAACGATATTATT 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 TCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCATGCTTTTGATGGAACGATATTATT 545

Qy        673 TTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTTTTTAGTAAGACCCGGAATGGAGA 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 TTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTTTTTAGTAAGACCCGGAATGGAGA 605

Qy        733 GGATGTGAGGATAACGATCACTTTAACAAA-TGAACTTCCACCTACATCACCAACTTGTT 791
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        606 GGATGTGAGGATAACGATCACTTTAACAAATTGAACTTCCACCTACATCACCAACTTGTT 665

Qy        792 TGCAGTTCTATAATATTATTTTCAGGAGGC-TTTTGAAAATCATGAATTTGC---AACAA 847
              |||||||||||||||||||||||||||||| |||||||||||||||||||    | ||
Db        666 TGCAGTTCTATAATATTATTTTCAGGAGGCTTTTTGAAAATCATGAATTTGCCAACAAAA 725

Qy        848 ATTGGACGAAATTATTATAACCCAAATGACCCAA-TTGATATTCCAAGTCACAGGTTGGT 906
              | ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db        726 TTGGGACGAAATTATTATAACCCAAATGACCCAATTTGATATTCCAAGTCACAGGTTGGT 785

Qy        907 GATTTGG-CCTGGCTTCACTACTTCCATCCTTC 938
              ||||||| |||||||||||| ||||||||||
Db        786 GATTTGGCCCTGGCTTCACTATTTCCATCCTTC 818


RESULT 2
BI560710
LOCUS       BI560710                779 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  603254744F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5296954 5',
            mRNA sequence.
ACCESSION   BI560710
VERSION     BI560710.1  GI:15448024
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 779)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11751  row: h  column: 11
            High quality sequence stop: 732.
FEATURES             Location/Qualifiers
     source          1. .779
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5296954"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      234 a    159 c    205 g    181 t
ORIGIN

  Query Match             21.0%;  Score 730.2;  DB 13;  Length 779;
  Best Local Similarity   99.2%;  Pred. No. 4e-145;
  Matches  754;  Conservative    0;  Mismatches    4;  Indels    2;  Gaps    2;

Qy        108 GGTCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAACAA 167
              || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         20 GGGCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAACAA 79

Qy        168 TGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGC 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         80 TGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGC 139

Qy        228 TGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGCCAC 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 TGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGCCAC 199

Qy        288 CAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAGCCA 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 CAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAGCCA 259
```

```
Qy        348 AGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAGAGAGGAG 407
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        260 AGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTAGCAGAGAGAGGAG 319

Qy        408 GTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGTTA 467
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 GTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGTTA 379

Qy        468 AAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGC 527
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 AAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGC 439

Qy        528 TGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGG 587
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 TGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGG 499

Qy        588 AAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTC 647
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 AAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTC 559

Qy        648 ATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAG 707
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 ATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAG 619

Qy        708 TTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATGAAC 767
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        620 TTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATGAAC 679

Qy        768 TTCCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTTGA 827
              ||||| |||||||| ||||||||||||||||||||||||||||||||||||||||| ||
Db        680 TTCCA-CTACATCAGCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTTGGA 738

Qy        828 AAATCA-TGAATTTGCAACAAATTGGACGAAATTATTATA 866
              |||||| ||||||||||||||||||||||||||||||||
Db        739 AAATCATTGAATTTGCAACAAATTGGACGAAATTATTATA 778
```

RESULT 3
BI463482
LOCUS BI463482 1092 bp mRNA linear EST 21-AUG-2001
DEFINITION 603203560F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5269311 5', mRNA sequence.
ACCESSION BI463482
VERSION BI463482.1 GI:15254125
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1092)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11679 row: h column: 16
High quality sequence stop: 668.

FEATURES Location/Qualifiers
source 1. .1092
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5269311"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 356 a 201 c 270 g 264 t 1 others
ORIGIN

Query Match 19.9%; Score 689.2; DB 13; Length 1092;
Best Local Similarity 83.5%; Pred. No. 2.1e-136;
Matches 903; Conservative 0; Mismatches 156; Indels 22; Gaps 10;

```
Qy        106 CAGGTCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAAC 165
              | ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         16 CGGGAGTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAAC 75

Qy        166 AATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCA 225
              ||    ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         76 AAGACTGGGAGAAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCA 135

Qy        226 G-CTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGC 284
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 GACTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGC 195

Qy        285 CACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAG 344
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 CACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAG 255

Qy        345 CCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAGAGAG 404
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        256 CCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTAGCAGAGAGAG 315

Qy        405 GAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATG 464
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 GAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATG 375

Qy        465 TTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCC 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 TTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCC 435

Qy        525 GGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGA 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGA 495

Qy        585 TGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGT 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 TGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGT 555

Qy        645 GTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTG 704
              |||||||||||||||||||||||||||||||||||||||||||||||||||| | |
Db        556 GTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAAGGTTACT 615

Qy        705 AAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCACTTTAACAAATG 764
              |  ||||||||||||||||||||||||||||||||||||||||||||||||  || |  ||
Db        616 GAAGTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATCAACTTTAACAAAT 675

Qy        765 AACTTCCACCTACATCACCAACTTGTTTGCAGTTCTATAATATTATTTTCAGGAGGCTTT 824
              |  | || ||||||   | |  |  ||| |  ||||||||||||||   |||||| ||
Db        676 GATCTTCAACTACAT--ACCAACTTGTTGAAAGTCTATAATATTATTTCAGGGAGGC-TT 732

Qy        825 TGAAAATCATGAATTTGCAACAAATTGGACGAAATTATTATA---ACCCAAATGACCCAA 881
              |||||||||||||||| ||||||||||||||||||||||||   || |||   ||| |
Db        733 TGAAAATCATGAATTTGAAACAAATTGGACGAAATTATTATAAGCACACAATGAACCGAG 792

Qy        882 TTGATATTCCAAGTCACAGGTTGGTG--ATTTGGCCTGGCTTCACTACTTCCA------T 933
              |||||||| ||||| ||||| ||| |  ||| | | ||| |  ||||||||||      |
Db        793 TTGATATTACAAGTAACAGGGTGGGGGAATTGGACATGGATATACTACTTCCAATTCCTT 852

Qy        934 CCTTCAGTATGAAAACAGCATCATGCT--CTGCACTGACGTTAGCCATAAAGTCC--TTC 989
              |      |    |||| |||||  |||         ||| ||||||||||    |  |
```

```
Db        853 ACAAATCTGATCAAACCGCATCTAGCTTCTGCGCGTGAGGTTAGCCATAANAGTCGTTCG 912

Qy        990 GAAGTGAGACTGTTTTGGATTTCATGTTC--AACTTTTATCATCAGACAGAAGAACATAA 1047
              ||||||   ||||| ||||||||| | ||  ||  || || ||      |||||||||
Db        913 AAAGTGAAGATGTTTGGGATTTCAGGATCCAAATGTTCATAATATAGACGAAGAACATTT 972

Qy       1048 ATTTCAAGAACAAGTTTCCAAAGAANTAATAGGTTTAGTTGTTCTTACCAAGTATAACAT 1107
              || | | |||| || || || |||| | |||  ||||| |||||   |||     |||
Db        973 ATGTTATGAACGAGGTTTCAGAGAAGT-ATAGAATTAGTGGTTCTAGACAACATATACAA 1031

Qy       1108 TAAGACATACAGAGTGGATGATATTGACTGGGACCAGAATCCCAAGAGCACCTTTAAGAA 1167
              | ||||||  |       | ||  ||| ||| | |  |||   | || |      |||||
Db       1032 TTAGACATCAAAGAGTAGTTATTATGAATGGCATCGAAATTAGATGATCTTTAGAAAGAA 1091

Qy       1168 A 1168
              |
Db       1092 A 1092
```

```
RESULT 4
BI459464
LOCUS       BI459464                882 bp    mRNA    linear   EST 21-AUG-2001
DEFINITION  603200456F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5266428 5',
            mRNA sequence.
ACCESSION   BI459464
VERSION     BI459464.1  GI:15250120
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 882)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11671  row: p  column: 13
            High quality sequence stop: 717.
FEATURES             Location/Qualifiers
     source          1. .882
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5266428"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      252 a    196 c    231 g    203 t
ORIGIN
```

```
  Query Match             19.4%;  Score 672.2;  DB 13;  Length 882;
  Best Local Similarity   91.5%;  Pred. No. 9e-133;
  Matches  789;  Conservative    0;  Mismatches   64;  Indels    9;  Gaps    7;

Qy          93 GCGGCCGCCCGGGCAGGTCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCA 152
               | ||| |   |    || ||||||||||||||||||||||||||||||||||||||||||
Db          22 GGGGCTGTTGGCCTCGGGCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCA 81
```

```
Qy         153 AGAAATAGAAAACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTC 212
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          82 AGAAATAGAAAACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTC 141

Qy         213 AGGAGACAGCGCAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTA 272
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         142 AGGAGACAGCGCAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTA 201

Qy         273 GGCCTCAGCCGCCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAG 332
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         202 GGCCTCAGCCGCCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAG 261

Qy         333 CAGGAGGAACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGN 392
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         262 CAGGAGGAACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGT 321

Qy         393 TAGCAGAGAGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGA 452
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         322 TAGCAGAGAGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGA 381

Qy         453 ACCTAGACCATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCA 512
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         382 ACCTAGACCATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCA 441

Qy         513 CTAACCATTTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACT 572
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         442 CTAACCATTTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACT 501

Qy         573 ATAACCCACTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATC 632
               |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         502 ATAA-CCACTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATC 560

Qy         633 TAATTGGAAAGTGTCATGCTTTTGATGG-AACGATATTATTTTTACCTAAAAGACTACAG 691
               |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db         561 TAATTGGAAAGTGTCATGCTTTTGATGGAAACGATATTATTTTTACCTAAAAGACTACAG 620

Qy         692 CAAAAGGTTACTGAAGTTTTTAGTAAGACCCGGAATGGAG-AGGATGTGAGGA-TAACGA 749
               |||||||||||||||||||||||||||| |||||||||| |||||||||||| || |||
Db         621 CAAAAGGTTACTGAAGTTTTTAGTAAGAACCGGAATGGAGAAGGATGTGAGGATTACCGA 680

Qy         750 TCACTTTAACAAATGAACTTCCACCTACAT--CACCAACTTGTTTGC--AGTTCTATAAT 805
               |||||||| || |||||||||||| || ||    | ||||||||||   ||||||||||
Db         681 TCACTTTACCACATGAACTTCCACGTAAATTCACCAAACTTGTTTGCCAGGTTCTATAAT 740

Qy         806 ATTATTTTCAGGAGGCTTTTGAAAATCATGAATTTGCAACAAATTGGACGAAATTATTAT 865
               |||||||   ||| ||||| || | ||  || ||||| || || ||| ||||| | ||||
Db         741 ATTATTTCACGGAAGCTTTGGACACTCCCGACTTTGCCACCAAATGGCCGAAAATTTTAT 800

Qy         866 AACCCAAATGACCCAATTGATATTCCAAGTCACAGGTTGGTGATTTGGCCTGGCTTCACT 925
               |  || ||||  |||||||||  |||| |||||| |||||||||| |||||| |  |
Db         801 ATACCGAATGGACCAATTGAGTTTCC-AGTCACCGGTTGGTGAATTGGCCGGATTCACTA 859

Qy         926 ACTTCCATCCTTCAGTATGAAA 947
               | | |||  | |  ||  || |
Db         860 AATGCCACGCCTAGGTTGGACA 881
```

```
RESULT 5
BG721550
LOCUS       BG721550                715 bp    mRNA    linear   EST 08-MAY-2001
DEFINITION  602695155F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4827445 5',
            mRNA sequence.
ACCESSION   BG721550
VERSION     BG721550.1  GI:14000737
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
```

```
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10743  row: m  column: 14
            High quality sequence stop: 715.
FEATURES             Location/Qualifiers
     source          1. .715
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4827445"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      212 a    160 c    189 g    154 t
ORIGIN

  Query Match             18.4%;  Score 637.2;  DB 12;  Length 715;
  Best Local Similarity   98.1%;  Pred. No. 2.6e-125;
  Matches  686;  Conservative    0;  Mismatches    9;  Indels    4;  Gaps    4;

Qy        109 GTCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAACAAT 168
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 GGCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAACAAT 76

Qy        169 GACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGCT 228
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 GACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGCT 136

Qy        229 GGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAG-CCGCCAC 287
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        137 GGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGACCGCCAC 196

Qy        288 CAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAGCCA 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 CAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAGCCA 256

Qy        348 AGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAGAGAGGAG 407
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        257 AGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTAGCAGAGAGAGGAG 316

Qy        408 GTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGTTA 467
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        317 GTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGTTA 376

Qy        468 AAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGC 527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        377 AAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGC 436

Qy        528 TGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGG 587
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        437 TGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGG 496

Qy        588 AAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTC 647
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        497 AAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTC 556

Qy        648 ATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAA-GGTTACTGAA 706
              |||||||||||||||||||||||||| |||||||||||||||||||| |||||||||
Db        557 ATGCTTTTGATGGAACGATATTATTTATACCTAAAAGACTACAGCAAAACGGTTACTGAA 616

Qy        707 GTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGAT-AACGATCACTTTAACAAA-TG 764
              |||| |||||||||||||||||||||||||||||||| | ||||||||||||||| ||
Db        617 GTTTATAGTAAGACCCGGAATGGAGAGGATGTGAGGATCACCGATCACTTTAACAAACTG 676

Qy        765 AACTTCCACCTACATCACCAACTTGTTTGCAGTTCTATA 803
              |||||||| ||||||||||||||||||  || ||||||||
Db        677 AACTTCCAGCTACATCACCAACTTGTCAGCCGTTCTATA 715


RESULT 6
BI463897
LOCUS       BI463897                 666 bp    mRNA    linear   EST 21-AUG-2001
DEFINITION  603202635F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5268465 5',
            mRNA sequence.
ACCESSION   BI463897
VERSION     BI463897.1  GI:15254553
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 666)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11677  row: e  column: 10
            High quality sequence stop: 666.
FEATURES             Location/Qualifiers
     source          1. .666
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5268465"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      194 a    144 c    187 g    141 t
ORIGIN

  Query Match             18.2%;  Score 633;  DB 13;  Length 666;
  Best Local Similarity   98.2%;  Pred. No. 2.1e-124;
  Matches  650;  Conservative    0;  Mismatches   11;  Indels    1;  Gaps    1;

Qy         93 GCGGCCGCCCGGGCAGGTCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCA 152
              | ||| |   |    || ||||||||||||||||||||||||||||||||||||||||||
Db          5 GGGGCTGTTGGCCTCGGGCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCA 64

Qy        153 AGAAATAGAAAACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTC 212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 AGAAATAGAAAACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTC 124

Qy        213 AGGAGACAGCGCAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTA 272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        125 AGGAGACAGCGCAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTA 184

Qy        273 GGCCTCAGCCGCCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAG 332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 GGCCTCAGCCGCCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAG 244

Qy        333 CAGGAGGAACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGN 392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 CAGGAGGAACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGT 304

Qy        393 TAGCAGAGAGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGA 452
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 TAGCAGAGAGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGA 364

Qy        453 ACCTAGACCATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCA 512
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 ACCTAGACCATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCA 424

Qy        513 CTAACCATTTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACT 572
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 CTAACCATTTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACT 484

Qy        573 ATAACCCACTGATGGAA-GCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGAT 631
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        485 ATAACCCACTGATGGAACGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGAT 544

Qy        632 CTAATTGGAAAGTGTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAG 691
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        545 CTAATTGGAAAGTGTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAG 604

Qy        692 CAAAAGGTTACTGAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATC 751
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        605 CAAAAGGTTACTGAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGATC 664

Qy        752 AC 753
              ||
Db        665 AC 666
```

RESULT 7
BI458664
LOCUS BI458664 731 bp mRNA linear EST 21-AUG-2001
DEFINITION 603199792F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5265838 5', mRNA sequence.
ACCESSION BI458664
VERSION BI458664.1 GI:15249320
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 731)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11670 row: g column: 23
High quality sequence stop: 671.

```
FEATURES             Location/Qualifiers
     source          1. .731
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5265838"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation). Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      214 a    157 c    201 g    159 t
ORIGIN

  Query Match           18.0%;  Score 626.2;  DB 13;  Length 731;
  Best Local Similarity 97.5%;  Pred. No. 5.8e-123;
  Matches  667;  Conservative    0;  Mismatches   14;  Indels    3;  Gaps    3;

Qy         93 GCGGCCGCCCGGGCAGGTCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCA 152
              | ||| |   |    || ||||||||||||||||||||||||||||||||||||||||||
Db         18 GGGGCTGTTGGCCTCGGGCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCA 77

Qy        153 AGAAATAGAAAACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTC 212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 AGAAATAGAAAACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTC 137

Qy        213 AGGAGACAGCGCAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTA 272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 AGGAGACAGCGCAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTA 197

Qy        273 GGCCTCAGCCGCCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAG 332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 GGCCTCAGCCGCCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAG 257

Qy        333 CAGGAGGAACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGN 392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 CAGGAGGAACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGT 317

Qy        393 TAGCAGAGAGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGA 452
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 TAGCAGAGAGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGA 377

Qy        453 ACCTAGACCATGTTAA-AGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGC 511
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        378 ACCTAGACCATGTTAACAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGC 437

Qy        512 ACTAACCATTTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGAC 571
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        438 ACTAACCATTTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGAC 497

Qy        572 TATAACCCACTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGAT 631
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        498 TATAACCCACTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGAT 557

Qy        632 CTAATTGGAAAGTGTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACA- 690
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        558 CTAATTGGAAAGTGTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAG 617

Qy        691 GCAAAAGGTTACT-GAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATAACGA 749
              |||||||||||||| |||||||||||||||||||||||||||||||||||||||||  ||
Db        618 GCAAAAGGTTACTGGAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTGAGGATACAGA 677

Qy        750 TCACTTTAACAAATGAACTTCCAC 773
              |||||||||||||||||||| ||
Db        678 TCACTTTAACAAATGAACTTCAAC 701
```

RESULT 8
AL705301
LOCUS AL705301 640 bp mRNA linear EST 22-MAR-2002
DEFINITION DKFZp686G1235_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone DKFZp686G1235 5', mRNA sequence.
ACCESSION AL705301
VERSION AL705301.1 GI:19688656

```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 640)
  AUTHORS   Ottenwaelder,B., Obermaier,B., Mewes,W., Mewes,H.W., Weil,B. and
            Wiemann,S.
  TITLE     EST (Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and Wiemann
            ,S.)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Ottenwaelder B
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by MediGenomix (Martinsried/Germany) within the cDNA
            sequencing consortium of the German Genome Project. No s1 sequence
            available.
            This clone (DKFZp686G1235) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .640
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686G1235"
                     /clone_lib="686 (synonym: hlcc3)"
                     /tissue_type="human skeletal muscle"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
BASE COUNT      190 a    139 c    174 g    137 t
ORIGIN

  Query Match             17.9%;  Score 621.4;  DB 9;  Length 640;
  Best Local Similarity   98.9%;  Pred. No. 6.2e-122;
  Matches  625;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        108 GGTCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAACAA 167
              || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          9 GGGCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAACAA 68

Qy        168 TGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGC 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 TGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGC 128

Qy        228 TGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGCCAC 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 TGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGCCAC 188

Qy        288 CAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAGCCA 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        189 CAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAGCCA 248

Qy        348 AGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAGAGAGGAG 407
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        249 AGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTAGCAGAGAGAGGAG 308

Qy        408 GTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGTTA 467
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        309 GTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGTTA 368

Qy        468 AAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGC 527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        369 AAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGC 428

Qy        528 TGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGG 587
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        429 TGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGG 488

Qy        588 AAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTC 647
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        489 AAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTC 548

Qy        648 ATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAG 707
              |||||||||||||||||||||||||||||||||||||||||||||| | ||||||||
Db        549 ATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAAGGTACTGAAG 608

Qy        708 TTTTTAGTAAGACCCGGAATGGAGAGGATGTG 739
              ||||||||||||||| ||||||| | ||||||
Db        609 TTTTTAGTAAGACCCCGAATGGAAAAGATGTG 640


RESULT 9
AL705072
LOCUS       AL705072                622 bp    mRNA    linear   EST 22-MAR-2002
DEFINITION  DKFZp686E0634_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
            DKFZp686E0634 5', mRNA sequence.
ACCESSION   AL705072
VERSION     AL705072.1  GI:19688427
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 622)
  AUTHORS   Ottenwaelder,B., Obermaier,B., Mewes,W., Mewes,H.W., Weil,B. and
            Wiemann,S.
  TITLE     EST (Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and Wiemann
            ,S.)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Ottenwaelder B
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by MediGenomix (Martinsried/Germany) within the cDNA
            sequencing consortium of the German Genome Project. No s1 sequence
            available.
            This clone (DKFZp686E0634) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .622
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686E0634"
                     /clone_lib="686 (synonym: hlcc3)"
                     /tissue_type="human skeletal muscle"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
BASE COUNT      187 a    134 c    168 g    133 t
ORIGIN

  Query Match             17.9%;  Score 621;  DB 9;  Length 622;
  Best Local Similarity   99.8%;  Pred. No. 7.6e-122;
  Matches  621;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        119 CAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAACAATGACTGGGAGA 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAACAATGACTGGGAGA 60

Qy        179 GCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGCTGGTGGGCTCC 238
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCAGCTGGTGGGCTCC 120

Qy        239 ACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGCCACCAGCAGAGGGG 298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGCCACCAGCAGAGGGG 180
```

```
Qy        299 GAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAGCCAAGTCACAAGGA 358
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAGCCAAGTCACAAGGA 240

Qy        359 CTCCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAGAGAGGAGGTCGTCGTAGA 418
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        241 CTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTAGCAGAGAGAGGAGGTCGTCGTAGA 300

Qy        419 GATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGTTAAAGAATCAAAA 478
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGTTAAAGAATCAAAA 360

Qy        479 ACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGACATCCCGT 538
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGACATCCCGT 420

Qy        539 CCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGGAAGCCAGAAGA 598
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGATGGAAGCCAGAAGA 480

Qy        599 CTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCATGCTTTTGAT 658
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTAATTGGAAAGTGTCATGCTTTTGAT 540

Qy        659 GGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTTTTTAGTAAG 718
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAGTTTTTAGTAAG 600

Qy        719 ACCCGGAATGGAGAGGATGTGA 740
              |||||||||||||||||||||||
Db        601 ACCCGGAATGGAGAGGATGTGA 622
```

RESULT 10
BQ435882

```
LOCUS       BQ435882                718 bp    mRNA    linear   EST 24-MAY-2002
DEFINITION  AGENCOURT_7897283 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:6103498
            5', mRNA sequence.
ACCESSION   BQ435882
VERSION     BQ435882.1  GI:21174947
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 718)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2339  row: n  column: 11
            High quality sequence stop: 518.
FEATURES             Location/Qualifiers
     source          1. .718
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6103498"
                     /clone_lib="NIH_MGC_82"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).   Average insert size
                     1.35 kb (range 0.9-4.0 kb).  14/15 colonies contained
                     inserts by PCR. This library was enriched for full-length
                     clones and was constructed by Clontech  Laboratories (Palo
                     Alto, CA)."
BASE COUNT      192 a    137 c    140 g    249 t
ORIGIN
```

```
  Query Match             16.7%;  Score 580.6;  DB 14;  Length 718;
  Best Local Similarity   98.3%;  Pred. No. 3e-113;
  Matches  640;  Conservative    0;  Mismatches    4;  Indels    7;  Gaps    5;

Qy       2504 GTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTATGAC 2563
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db             5 GTGAGCCAGGCTGTGAGAAGTGGTAGTGTTTCTCCCACACATTACAATGTCATCTATGAC 64

Qy       2564 AACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATCTAT 2623
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            65 AACAGCGGCCTGAAGCCAGACCACATACAGCGCTTGACCTACAAGCTGTGCCACATCTAT 124

Qy       2624 TACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCT 2683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           125 TACAACTGGCCAGGTGTCATTCGTGTTCCTGCTCCTTGCCAGTACGCCCACAAGCTGGCT 184

Qy       2684 TTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTAC 2743
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           185 TTTCTTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAACCGCCTTTAC 244

Qy       2744 TACCTCTAACCTGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTT 2803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           245 TACCTCTAACCTGCAGAAGACGATGCAGCCGCTTTTCTTTTTGAAATGACTTTGGGATTT 304

Qy       2804 TTTTAAGCTTTTATTTACTTTTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGG 2863
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db           305 TTTTAAGCTTTTATTTAC-TTTTTTTTAACTGTTATCTTTCTGGATGAAACTTGGGAAGG 363

Qy       2864 GGATTAGGAGATCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTAT 2923
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           364 GGATTAGGAGATCTAGCATTTTATTTCTAGCATTGCTATTCACCGGCTTCCTTATTTTAT 423

Qy       2924 ATGTAAAAATTAAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGC-A 2982
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db           424 ATGTAAAAATTAAGATTTTATATTTTATCTTCTTGTTTCTCATAGATATTTTGTGAGCAA 483

Qy       2983 TTTTTTTGTTTATTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTC 3042
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           484 TTTTTTTGTTTATTTTGAAGAAATGTGGATAAGATACTTGGTAGTATAAAACAGACTCTC 543

Qy       3043 TGAGAGTATTTGAAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGC-AAG 3101
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||| 
Db           544 TGAGAGTATTTGAAATGTGTTTGGAGATTTACTTAAACGTACTTTCAGGAGTGAGCAAAG 603

Qy       3102 TCCTACTTA-TAAACCTATATTAACTTT---ATTTTTGAGATACCTGTTTT 3148
              ||||||||| |||||||||||||||| ||    ||||||| ||||||| |||
Db           604 TCCTACTTATTAAACCTATATTAACCTTTAATTTTTTGAAATACCTGGTTT 654
```

RESULT 11
BF666315

```
LOCUS       BF666315                874 bp    mRNA    linear   EST 21-DEC-2000
DEFINITION  602119233F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4276136 5',
            mRNA sequence.
ACCESSION   BF666315
VERSION     BF666315.1  GI:11940210
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 874)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
```

```
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: ATCC
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLCM1096  row: b  column: 09
           High quality sequence stop: 594.
FEATURES             Location/Qualifiers
     source          1. .874
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4276136"
                     /clone_lib="NIH_MGC_56"
                     /tissue_type="primitive neuroectoderm"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA.  5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.65 kb (range 0.9-4.0 kb).  15/15 colonies
                     contained  inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      200 a    218 c    258 g    198 t
ORIGIN

  Query Match             16.5%;  Score 574.6;  DB 12;  Length 874;
  Best Local Similarity   93.8%;  Pred. No. 5.6e-112;
  Matches  653;  Conservative    0;  Mismatches   35;  Indels    8;  Gaps    5;

Qy       1757 ATGATTGAAGTGGATGACAGAACTGAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACA 1816
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          8 AGGATTGAAGTGGATGACAGAACTGAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACA 67

Qy       1817 GCAGACACCCAGATAGTTGTCTGTCTGTTGTCAAGTAATCGGAAGGACAAATACGATGCT 1876
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 GCAGACACCCAGATAGTTGTCTGTCTGTTGTCAAGTAATCGGAAGGACAAATACGATGCT 127

Qy       1877 ATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACC 1936
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        128 ATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACC 187

Qy       1937 TTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAAAGATTGCCCTACAGATGAACTGC 1996
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        188 TTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAAAGATTGCCCTACAGATGAACTGC 247

Qy       1997 AAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGC 2056
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        248 AAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGC 307

Qy       2057 ATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGC 2116
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        308 ATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTGCCAGC 367

Qy       2117 ATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCATATTTCAGGATAGAGGACAGGAG 2176
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        368 ATCAATGAAGGGATGACCCGCTGGTTCTCACGCTGCATATTTCAGGATAGAGGACAGGAG 427

Qy       2177 CTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAAT 2236
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        428 CTGGTAGATGGGCTCAAAGTCTGCCTGCAAGCGG-TCTGAGGGCTTGGAATAGCTGCAAT 486

Qy       2237 GAGTACATGCCCAGCCGGATCATCGTGTACCGCGATGGCGTAGGAGACGGCCAGCTGAAA 2296
              |||||||||||||||||||||||||||||||||||||||||||||||||||     |||||
Db        487 GAGTACATGCCCAGCCGGATCATCGTGTACCGCGATGGCGTAGGAGACGGC--AGTGAAA 544

Qy       2297 ACACTGGTGAACTACGAAGTGCCACAGTTTTTGGATTGTCTAAAATCCATTGGTAGAGGT 2356
              |||||||||||||||||||||||||||||| ||||||| |||||||||||| | |||||||
Db        545 ACACTGGTGAACTACGAAGTGCCACAG-TTTTGGATAGTCTAAAATCCATGGTTAGAGGT 603

Qy       2357 TACAACCCTAGANTAACGGTAATTGTGGTGAAGAAAAGAGTGAACACCAGATTTTTTGCT 2416
              | |||||||||| |||||||| | |||||| |  |  ||||| || ||    ||||
Db        604 T-CAACCCTAGACTAACGGTACTCGTGGTGGACGACCGAGTGGACCCCCAGATTTT---G 659

Qy       2417 CAGTCTGGAGGAAGACTTCAGAATCCACTTCCTGGA 2452
              |      | ||| |||||| ||   |  ||||||||
Db        660 CTCCGCCGCGGACGACTTCGGACATCCTTTCCTGGA 695


RESULT 12
BG718168
LOCUS       BG718168                581 bp    mRNA    linear   EST 08-MAY-2001
DEFINITION  602696127F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4828038 5',
            mRNA sequence.
ACCESSION   BG718168
VERSION     BG718168.1  GI:13997355
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 581)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10745  row: f  column: 07
            High quality sequence stop: 578.
FEATURES             Location/Qualifiers
     source          1. .581
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4828038"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      164 a    132 c    165 g    120 t
ORIGIN

  Query Match             16.1%;  Score 560.6;  DB 12;  Length 581;
  Best Local Similarity   98.3%;  Pred. No. 5.6e-109;
  Matches  566;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy         95 GGCCGCCCGGGCAGGTCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAG 154
              ||| |   |    || ||||||||||||||||||||||||||||||||||||||||||||
Db          5 GGCTGTTGGCCTCGGGCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAG 64

Qy        155 AAATAGAAAACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAG 214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 AAATAGAAAACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAG 124

Qy        215 GAGACAGCGCAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGG 274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        125 GAGACAGCGCAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGG 184

Qy        275 CCTCAGCCGCCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCA 334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 CCTCAGCCGCCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCA 244

Qy        335 GGAGGAACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGNTA 394
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        245 GGAGGAACAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTA 304

Qy        395 GCAGAGAGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAAC 454
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 GCAGAGAGAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAAC 364

Qy        455 CTAGACCATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACT 514
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 CTAGACCATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACT 424

Qy        515 AACCATTTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTAT 574
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 AACCATTTCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTAT 484

Qy        575 AACCCACTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTA 634
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 AACCCACTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTA 544

Qy        635 ATTGGAAAGTGTCATGCTTTTGATGGAACGATATTA 670
              |||||||||||||||||||||||||||||||||||
Db        545 ATTGGAAAGTGTCATGCTTTTGATGGAACGATATTA 580


RESULT 13
BI463615
LOCUS       BI463615                539 bp    mRNA    linear   EST 21-AUG-2001
DEFINITION  603206270F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5272073 5',
            mRNA sequence.
ACCESSION   BI463615
VERSION     BI463615.1  GI:15254271
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11686  row: k  column: 18
            High quality sequence stop: 539.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5272073"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      153 a    127 c    154 g    105 t
ORIGIN

  Query Match             15.1%;  Score 523.2;  DB 13;  Length 539;
  Best Local Similarity   99.2%;  Pred. No. 5.2e-101;
  Matches  525;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        106 CAGGTCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAAC 165
              | ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 CGGGCTTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAAAAC 70

Qy        166 AATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCA 225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         71 AATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGCGCA 130

Qy        226 GCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGCC 285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 GCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCCGCC 190

Qy        286 ACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAGC 345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 ACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACAGCAGGAGGAACAGC 250

Qy        346 CAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAGAGAGG 405
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        251 CAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTAGCAGAGAGAGG 310

Qy        406 AGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGT 465
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        311 AGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACCATGT 370

Qy        466 TAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCG 525
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        371 TAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATTTCCG 430

Qy        526 GCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGAT 585
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        431 GCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCACTGAT 490

Qy        586 GGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTA 634
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        491 GGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAACACGAAGATCTA 539


RESULT 14
BI459142
LOCUS       BI459142                744 bp    mRNA    linear   EST 21-AUG-2001
DEFINITION  603199321F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5265484 5',
            mRNA sequence.
ACCESSION   BI459142
VERSION     BI459142.1  GI:15249798
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 744)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11669  row: i  column: 05
```

```
          High quality sequence stop: 667.
FEATURES             Location/Qualifiers
     source          1. .744
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5265484"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      226 a    155 c    207 g    156 t
ORIGIN

  Query Match           15.0%;  Score 521;  DB 13;  Length 744;
  Best Local Similarity   92.6%;  Pred. No. 1.5e-100;
  Matches  624;  Conservative    0;  Mismatches   41;  Indels    9;  Gaps    7;

Qy        103 GGGCAGGTCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAA 162
              ||   || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         29 GGCTCGGGCTGAGGTGCAAGGACCAGGACTAGGGCGAGGGCAGCGGTCCAAGAAATAGAA 88

Qy        163 AACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGC 222
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         89 AACAATGACTGGGAGAGCCCGAGCCAGAGCCAGAGGAAGGGCCCGCGGTCAGGAGACAGC 148

Qy        223 GCAGCTGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCC 282
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        149 GCAG-TGGTGGGCTCCACTGCCAGTCAGCAACCTGGTTATATTCAGCCTAGGCCTCAGCC 207

Qy        283 GCCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAA-CAGCAGGAGGAA 341
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        208 GCCACCAGCAGAGGGGGAATTATTTGGCCGTGGACGGCAGAGAGGAACCAGCAGGAGGAA 267

Qy        342 CAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGNTAGCAGAGA 401
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        268 CAGCCAAGTCACAAGGACTCCAGATATCTGCTGGATTTCAGGAGTTATCGTTAGCAGAGA 327

Qy        402 GAGGAGGTCGTCGTAGAGATTTTCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACC 461
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        328 GAGGAGGTCGTCGTAGAGATTTCCATGATCTTGGTGTGAATACAAGGCAGAACCTAGACC 387

Qy        462 ATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATT 521
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        388 ATGTTAAAGAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTAACCATT 447

Qy        522 TCCGGCTGACATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCC-- 579
              |||||     ||||||||||||||||||||||||||||||||||||||||||||||
Db        448 TCCGGTGACAATCCCGTCCCCAGTGGGCCTTATATCAGTATCACATTGACTATAACCCCA 507

Qy        580 ACTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAA--CACGAAGATCTAATT 637
              ||||||||||||||||||||||||||||||||||||||||||   |||||      | |
Db        508 CGTGATGGAAGCCAGAAGACTCCGTTCAGCTCTTCTTTTTCAAGCAACGAAAGATCTAAT 567

Qy        638 GGAAAGTGTCATGCTTTTGATGGAACGATATTATTTTTACCTAAAAGACTACAGCAAAAG 697
              |||||||||||||| ||||||||||||||||||||||||||||      |||||
Db        568 GGAAAGTGTCATGC-TTTGATGGAACGATATTATTTTTACCTAAAAAGACTACAGAAAAG 626

Qy        698 GTTACTGAAGTTTTTAGTAAGACCCGGAATGGAGAGGATGTG-AGGATAACGATCACTTT 756
              |||||||||| ||| ||||||| ||||||||||||||||||| ||||||| || || |||
Db        627 GTTACTGAAG-TTTGAGTAAGAACCGGAATGGAGAGGATGTGAAGGATAAGGACCAATTT 685

Qy        757 AACAAATGAACTTC 770
              ||| ||   | | |
Db        686 AACGAAATGAATGC 699
```

```
RESULT 15
BF080918
LOCUS       BF080918                577 bp    mRNA    linear   EST 18-OCT-2000
DEFINITION  233684 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF080918
VERSION     BF080918.1  GI:10874739
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 577)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 53  row: N  column: 21
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .577
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      198 a     98 c    132 g    149 t
ORIGIN

  Query Match           14.8%;  Score 515.2;  DB 12;  Length 577;
  Best Local Similarity   93.4%;  Pred. No. 2.6e-99;
  Matches  538;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       1556 CAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCCAAAGAAACAAGA 1615
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db          2 CAAGGTGGAAAAACATTTGATTACAATCCACAATTTGCAGATTGGTCAAAAGAAACAAGA 61

Qy       1616 GGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACGCGAAGA 1675
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db         62 GGTGCACCATTAATTAGTGTTAAGCCACTAGATAACTGGCTGTTGATCTATACTCGAAGA 121

Qy       1676 AATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGC 1735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 AATTATGAAGCAGCCAATTCATTGATACAAAATCTATTTAAAGTTACACCAGCCATGGGC 181

Qy       1736 ATGCAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGAACTGAAGCCTACTTAAGA 1795
              || ||||||| |||||||||||||||||||||||| ||||||||||||| ||||||
Db        182 ATACAAATGAAAAAAGCAATAATGATTGAAGTGGATGATAGAACTGAAGCCTATTTAAGA 241

Qy       1796 GTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCTGTTGTCAAGTAAT 1855
              |||||||||||||||| ||| | || ||||||||||||| ||||||||||||||||||
Db        242 GTCTTACAGCAAAAGGTTACATCGGATACCCAGATAGTTGTTTGTCTGTTGTCAAGTAAT 301

Qy       1856 CGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGT 1915
              || |||||||||| |||||||||||||||| | ||||||||||||||| ||||||
Db        302 CGAAAGGACAAATATGATGCTATTAAAAAATACTTATGTACAGATTGCCCTACTCCAAGT 361
```

```
Qy       1916 CAGTGTGTGGTGGCCCGAACCTTAGGCAAACAGCAAACTGTCATGGCCATTGCTACAAAG 1975
              ||||||||||||||||||||||||||||||||| || |||||||||||||||| ||||||
Db        362 CAGTGTGTGGTGGCCCGAACCTTAGGCAAACAACAGACTGTCATGGCCATTGCGACAAAG 421

Qy       1976 ATTGCCCTACAGATGAACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTG 2035
              || ||||| |||||||||||||||||||||||||||| |||||||| || || || ||
Db        422 ATCGCCCTGCAGATGAACTGCAAGATGGGAGGAGAGCTTTGGAGGGTTGATATGCCTCTA 481

Qy       2036 AAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCA 2095
              ||||| | |||||| |||||||| |||||||| ||||| |  |||||||| |||||||||
Db        482 AAGCTAGCGATGATAGTTGGCATTGATTGTTATCATGATACCACAGCTGGACGGAGGTCA 541

Qy       2096 ATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATG 2131
              || |||||||||||||| ||||||||||||||||||
Db        542 ATTGCAGGATTTGTTGCAAGCATCAATGAAGGGATG 577
```

Search completed: July 21, 2003, 22:53:06
Job time : 4562.56 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2003, 19:01:27 ; Search time 30.1988 Seconds
(without alignments)
3799.123 Million cell updates/sec

Title: US-09-873-737A-6
Perfect score: 4534
Sequence: 1 MTGRARARARGRARGQETAQ.....VGQSIHREPNLSLSNRLYYL 861

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4528	99.9	861	21	AAY90235	Human piwi protein
2	4286.5	94.5	862	21	AAY90234	Mouse piwi protein
3	1464	32.3	866	22	ABB62102	Drosophila melanog
4	1417	31.3	843	22	ABB62084	Drosophila melanog
5	1406	31.0	843	21	AAY90233	Drosophila piwi pr
6	1365.5	30.1	580	22	AAU07866	Polypeptide sequen
7	1220	26.9	530	22	AAB94209	Human protein sequ
8	959	21.2	371	22	AAB92722	Human protein sequ
9	652	14.4	130	21	AAB58977	Breast and ovarian
10	559	12.3	341	22	ABG05957	Novel human diagno
11	521	11.5	983	21	AAG51562	Arabidopsis thalia
12	521	11.5	988	21	AAG51561	Arabidopsis thalia
13	517	11.4	1048	21	AAG42118	Arabidopsis thalia
14	514.5	11.3	906	21	AAG51563	Arabidopsis thalia
15	510	11.2	1064	21	AAG52593	Arabidopsis thalia
16	498	11.0	868	21	AAG42119	Arabidopsis thalia
17	493	10.9	950	22	ABB67046	Drosophila melanog
18	491	10.8	884	21	AAG52594	Arabidopsis thalia
19	490	10.8	984	22	ABB62573	Drosophila melanog
20	490	10.8	984	22	ABB67045	Drosophila melanog
21	486	10.7	860	22	AAM80292	Human protein SEQ
22	486	10.7	860	22	AAB94291	Human protein sequ
23	473.5	10.4	820	21	AAG42120	Arabidopsis thalia
24	473.5	10.4	836	21	AAG52595	Arabidopsis thalia
25	462.5	10.2	782	22	AAB93139	Human protein sequ
26	436	9.6	1145	22	AAE09770	Drosophila melanog
27	421	9.3	962	22	ABB63278	Drosophila melanog
28	368	8.1	938	22	AAB82553	Neurospora crassa
29	340.5	7.5	199	22	ABG05956	Novel human diagno
30	319.5	7.0	326	21	AAG52610	Arabidopsis thalia
31	306	6.7	1020	22	AAU01856	C. elegans RNA int
32	301	6.6	140	22	ABG12914	Novel human diagno
33	295.5	6.5	406	21	AAG54509	Zea mays protein f
34	290	6.4	227	21	AAG52611	Arabidopsis thalia
35	287	6.3	252	22	AAB97112	Translation initia
36	282	6.2	108	22	AAO04945	Human polypeptide
37	249.5	5.5	352	21	AAG54511	Zea mays protein f
38	249.5	5.5	353	21	AAG54510	Zea mays protein f
39	217	4.8	232	22	AAM85465	Human immune/haema
40	210.5	4.6	71	23	ABG40429	Human peptide enco
41	202	4.5	747	21	AAG25621	Arabidopsis thalia
42	202	4.5	749	21	AAG25620	Arabidopsis thalia
43	183	4.0	567	21	AAG25622	Arabidopsis thalia
44	145	3.2	55	22	ABB37543	Peptide #5049 enco
45	145	3.2	55	22	ABB22837	Protein #4836 enco

ALIGNMENTS

RESULT 1
AAY90235
ID AAY90235 standard; Protein; 861 AA.
XX
AC AAY90235;
XX
DT 29-AUG-2000 (first entry)
XX
DE Human piwi protein, designated hiwi.
XX
KW Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW anaemia; immunodeficiency; male infertility; human.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 76
FT /label= Leu, Ile
FT /note= "encoded by NTA"
FT Misc-difference 303
FT /label= Leu, Ile
FT /note= "encoded by NTA"
FT Misc-difference 735
FT /label= Leu, Ile
FT /note= "encoded by NTA"
XX
PN WO200032039-A1.
XX
PD 08-JUN-2000.
XX
PF 03-DEC-1999; 99WO-US28764.
XX
PR 04-DEC-1998; 98US-0110901.

```
XX
PA   (UYDU-) UNIV DUKE.
XX
PI   Lin H;
XX
DR   WPI; 2000-412085/35.
DR   N-PSDB; AAA07588.
XX
PT   Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT   therapy of diseases such as cancer and in various research and
PT   diagnostic applications -
XX
PS   Claim 4; Page 189-194; 201pp; English.
XX
CC   This sequence represents the human piwi family protein, designated
CC   hiwi. The piwi family nucleic acids and polypeptides are used in gene
CC   therapy of diseases such as cancer and also in various research and
CC   diagnostic applications. The sequences can also be used to treat
CC   tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ   Sequence   861 AA;

  Query Match             99.9%;  Score 4528;  DB 21;  Length 861;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  861;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRGTAGGTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRGTAGGTA 60

Qy         61 KSQGLQISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFR 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 KSQGLQISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFR 120

Qy        121 LTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTE 180

Qy        181 VFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPID 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPID 240

Qy        241 IPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVS 300

Qy        301 KEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQP 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQP 360

Qy        361 VLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQRE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQRE 420

Qy        421 VGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 VGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSK 480

Qy        481 ETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDRTEA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDRTEA 540

Qy        541 YLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAI 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAI 600

Qy        601 ATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFS 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFS 660

Qy        661 RCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQF 720

Qy        721 LDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIV 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 LDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIV 780

Qy        781 SQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 SQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAF 840

Qy        841 LVGQSIHREPNLSLSNRLYYL 861
              |||||||||||||||||||||
Db        841 LVGQSIHREPNLSLSNRLYYL 861


RESULT 2
AAY90234
ID   AAY90234 standard; Protein; 862 AA.
XX
AC   AAY90234;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Mouse piwi protein, designated miwi.
XX
KW   Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW   anaemia; immunodeficiency; male infertility; mouse; ds.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 90
FT                   /label= Leu, Ile
FT                   /note= "encoded by NTT"
FT   Misc-difference 216
FT                   /note= "unspecified amino acid; encoded by NTC"
FT   Misc-difference 383
FT                   /label= Leu, Ile
FT                   /note= "encoded by NTC"
FT   Misc-difference 816
FT                   /label= Leu, Ile
FT                   /note= "encoded by NTC"
XX
PN   WO200032039-A1.
XX
PD   08-JUN-2000.
XX
PF   03-DEC-1999;   99WO-US28764.
XX
PR   04-DEC-1998;   98US-0110901.
XX
PA   (UYDU-) UNIV DUKE.
XX
PI   Lin H;
XX
DR   WPI; 2000-412085/35.
DR   N-PSDB; AAA07587.
XX
PT   Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT   therapy of diseases such as cancer and in various research and
PT   diagnostic applications -
XX
PS   Claim 4; Page 180-185; 201pp; English.
XX
CC   This sequence represents the mouse piwi family protein, designated
CC   miwi. The piwi family nucleic acids and polypeptides are used in gene
CC   therapy of diseases such as cancer and also in various research and
CC   diagnostic applications. The sequences can also be used to treat
CC   tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ   Sequence   862 AA;

  Query Match             94.5%;  Score 4286.5;  DB 21;  Length 862;
  Best Local Similarity   93.7%;  Pred. No. 0;
```

Matches 808; Conservative 22; Mismatches 31; Indels 1; Gaps 1;

```
Qy          1 MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRG-TAGGT 59
              |||||||||||||||||| | ||: |||||||| |||| | ||:| |||||||   | |
Db          1 MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEGDLVGRGRQRGMVVGAT 60

Qy         60 AKSQGLQISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHF 119
              :||| |||||||||||| |||||||||||| |||||||||||||||||||||:||||||
Db         61 SKSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHF 120

Qy        120 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVT 179
              |||||||||||||||||||||||||||||||||||||:|||||||||||||||| |||
Db        121 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVT 180

Qy        180 EVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPI 239
              ||||:||||| |||||||||||||||||||||| |||||||||||||||||||||:|||
Db        181 EVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPI 240

Qy        240 DIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQV 299
              |||:|||||||||||||||:|||||||||||||||||||||||| | ||||||||||
Db        241 DIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQV 300

Qy        300 SKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQ 359
              ||| |||:|||||| |||||||||||||||||||||| ||||||||||||||||||||
Db        301 SKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQ 360

Qy        360 PVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 419
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        361 PVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 420

Qy        420 EVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWS 479
              ||||||||||:|||||||||||||||||||||||||||:||||||||||||||||||||
Db        421 EVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWS 480

Qy        480 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDRTE 539
              |||||||||||||||||||||||||||||||||||||||||:||:||||||||||||
Db        481 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTE 540

Qy        540 AYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 599
              |||| ||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 600

Qy        600 IATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWF 659
              |||||||||||||||||||: ||| |||||||||| ||||||||||||||||||||||||
Db        601 IATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWF 660

Qy        660 SRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQ 719
              |||:|||||||||||||||||||||: |||||||:|||||||||||||||||||||||||
Db        661 SRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQ 720

Qy        720 FLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 779
              |||||||:||||||| |||||||||| ||||||||||||||||||||||||||||||||
Db        721 FLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 780

Qy        780 VSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLA 839
              |||||||||||||||||:|||||||||||||| ||:|||||||||||||||||||||||
Db        781 VSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLA 840

Qy        840 FLVGQSIHREPNLSLSNRLYYL 861
              ||||||||||||||||||||||
Db        841 FLVGQSIHREPNLSLSNRLYYL 862
```

RESULT 3
ABB62102
ID ABB62102 standard; Protein; 866 AA.
XX
AC ABB62102;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 13098.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL06205.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 13098; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 866 AA;

Query Match 32.3%; Score 1464; DB 22; Length 866;
Best Local Similarity 36.6%; Pred. No. 2.6e-131;
Matches 323; Conservative 165; Mismatches 340; Indels 54; Gaps 17;

```
Qy          9 ARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRG--TAGGTAKSQGLQ 66
              |||| ||          ::|  ::      |: |:        |   :|| |
Db         10 ARGRGRG-----------RKPNNVEANRGFAPSLGQKSDPSHSEGNQASGGNGGGGDAQ 57

Qy         67 ISAGFQELSXA----------ERGG--RRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRL 114
              :    :: | :           ||    ||   ||  :    :      || | |  : :
Db         58 VGPSIEKSSLSAVQMHKSEGDPRGSVRGRRLITDLVYSRPPGM----TSKKGVVGTHITV 113

Qy        115 STNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLP--- 171
              |:|::  || | :||| :|: | :|| ||| : |::|: ::|  : |||| :|
Db        114 QANYFKVLKRPNWTIYQYRVDFTPDVEATRLRRSFLYEHKGILGG-YIFDGTNMFCINQF 172

Qy        172 KRLQQK--VTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIG 229
              | :|    | |: :|:| ||:: | |     :  |     |  |:| || :: ::|: :
Db        173 KAVQDSPYVLELVTKSRAGENIEIKIKAVGSVQSTDAEQFQVLNLILRRAMEGLDLKLVS 232

Qy        230 RNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQ 289
              | ||:|   |:: : |: :|||: ||| |:|| |:||::: |||:|:||: : : :
Db        233 RYYYDPQAKINLENFRMQLWPGYQTSIRQHENDILLCSEICHKVMRTETLYNILSDAIRD 292

Qy        290 TEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQ 349
              :::  :|    :  :|:|:|| || ||||:||:|:   |   ||  ||  :|:::||:|:
Db        293 SDD--YQSTFKRAVMGMVILTDYNNKTYRIDDVDFQSTPLCKFKTNDGE-ISYVDYYKKR 349

Qy        350 YNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVH 409
              ||  | |||||:::|:|  :   ||      |:||||   ||:|| || ||  :: :: |
Db        350 YNIIIRDLKQPLVMSRPTDKNIRGGN-DQAIMIIPELARATGMTDAMRADFRTLRAMSEH 408
```

```
Qy        410 TRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTF 469
Db        409 TRLNPDRRIERLRMFNKRLKSCKQSVETLKSWNIELDSALVEIPARVLPPEKILFGNQKI 468

Qy        470 DYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKA 529
Db        469 FVCDARADWTNEFRTCSMFKNVHINRWYVITPSRNLRETQEFVQMCIRTASSMKMNICNP 528

Qy        530 IMIEV-DDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVA 588
Db        529 IYEEIPDDRNGTYSQAIDNAAANDPQIVMVVMRSPNEEKYSCIKKRTCVDRPVPSQVVTL 588

Qy        589 RTLGKQQT----VMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSI 644
Db        589 KVIAPRQQKPTGLMSIATKVVIQMNAKLMGAPWQVVIPLHGLMTVGFDVCHSPKNKNKSY 648

Qy        645 AGFVASINEGMT-RWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDG 703
Db        649 GAFVATMDQKESFRYFSTVNEHIKGQELSEQMSVNMACALRSYQEQHRSLPERILFFRDG 708

Qy        704 VGDGQLKTLVNYEVPQFLDCL----KSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNP 759
Db        709 VGDGQLYQVVNSEVNTLKDRLDEIYKSAGKQEGCRMTFIIVSKRINSRYF--TGHR--NP 764

Qy        760 LPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIY 819
Db        765 VPGTVVDDVITLPERYDFFLVSQAVRIGTVSPTSYNVISDNMGLNADKLQMLSYKMTHMY 824

Qy        820 YNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 861
Db        825 YNYSGTIRVPAVCHYAHKLAFLVAESINRAPSAGLQNQLYFL 866
```

RESULT 4
ABB62084
ID ABB62084 standard; Protein; 843 AA.
XX
AC ABB62084;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 13044.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL06187.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 13044; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 843 AA;

Query Match 31.3%; Score 1417; DB 22; Length 843;
Best Local Similarity 35.9%; Pred. No. 8.6e-127;
Matches 309; Conservative 170; Mismatches 313; Indels 68; Gaps 19;

```
Qy         48 GRGRQR-------GTAGGTAKSQGLQISAG--------------------FQELSXAERG 80
Db          6 GRGRRRPLNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG 65

Qy         81 GRRR----DFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDY 136
Db         66 PTERKPWGDQYDY-LNTR---PAELVSKKGTDGVPVMLQTNFFRLKTKPEWRIVHYHVEF 121

Qy        137 NPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITIT 196
Db        122 EPSIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIK 181

Qy        197 LTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSI 256
Db        182 FVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSI 241

Qy        257 LQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKT 316
Db        242 RQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRT 299

Qy        317 YRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTL 376
Db        300 YRINDVDFGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNA 357

Qy        377 PGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQ- 435
Db        358 SELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPE 414

Qy        436 --RELRDWGLSFDSNLLSFSGRILQTEKI--HQGGKTFDYNPQFADWSKETRGAPLISVK 491
Db        415 SVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTP 471

Qy        492 P--LDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRK-AIMIEVDDRTEAYLRVLQQK 548
Db        472 SDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDC 531

Qy        549 VTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQM 608
Db        532 VRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQL 590

Qy        609 NCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASIN-EGMTRWFSRCIFQDR 667
Db        591 NCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSA 650

Qy        668 GQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSI 727
Db        651 FDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKT- 709

Qy        728 GRGY------NPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVS 781
Db        710 --EYARVQLSPPQLAYIVVTRSMNTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVS 763

Qy        782 QAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFL 841
Db        764 QQVRQGTVSPTSYNVLYSSMGLSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATL 823

Qy        842 VGQSIHREPNLSLSNRLYYL 861
```

```
Db        824 VGTNLHSIPQNALEKKFYYL 843


RESULT 5
AAY90233
ID   AAY90233 standard; Protein; 843 AA.
XX
AC   AAY90233;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Drosophila piwi protein.
XX
KW   Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW   anaemia; immunodeficiency; male infertility; Drosophila.
XX
OS   Drosophila sp.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 13
FT                   /label= Leu, Ile
FT                   /note= "encoded by NTT"
FT   Misc-difference 106
FT                   /label= Leu, Ile
FT                   /note= "encoded by NTA"
FT   Misc-difference 785
FT                   /label= Leu, Ile
FT                   /note= "encoded by NTC"
XX
PN   WO200032039-A1.
XX
PD   08-JUN-2000.
XX
PF   03-DEC-1999;   99WO-US28764.
XX
PR   04-DEC-1998;   98US-0110901.
XX
PA   (UYDU-) UNIV DUKE.
XX
PI   Lin H;
XX
DR   WPI; 2000-412085/35.
DR   N-PSDB; AAA07586.
XX
PT   Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT   therapy of diseases such as cancer and in various research and
PT   diagnostic applications -
XX
PS   Claim 4; Page 171-175; 201pp; English.
XX
CC   This sequence represents the Drosophila piwi family protein, designated
CC   piwi. The piwi family nucleic acids and polypeptides are used in gene
CC   therapy of diseases such as cancer and also in various research and
CC   diagnostic applications. The sequences can also be used to treat
CC   tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ   Sequence   843 AA;

  Query Match             31.0%;  Score 1406;  DB 21;  Length 843;
  Best Local Similarity   35.8%;  Pred. No. 9.8e-126;
  Matches  308;  Conservative  172;  Mismatches  312;  Indels   68;  Gaps   20;

Qy         48 GRGRQR-------GTAGGTAKSQGLQISAG--------------------FQELSXAERG 80
              ||||:|        |: |:     :::  |                     :|    | |
Db          6 GRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG 65

Qy         81 GRRR----DFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDY 136
                 |    | :|  :|||  :: |  || |: |: | | || ||  ::|:| :  ||:::
Db         66 PPERKPWGDQYDY-LNTRP-VELV--SKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEF 121

Qy        137 NPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITIT 196
                | :|  |:|  :|  | :|:|  : |||  ||  :: :|::| :  |::   : :|:|
Db        122 VPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIK 181

Qy        197 LTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSI 256
                 :    |  ||  |:| || :| :||: :||| ::|   |:|   :: :|||: |||
Db        182 FVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSI 241

Qy        257 LQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKT 316
               |:|  |:| |:::|||:|:||: | |    |    |  |::|    : |:||| || :|
Db        242 RQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRT 299

Qy        317 YRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTL 376
              ||::|:|: | |||||    |  :||:|||  :||  | |  ||:|:|:  | :
Db        300 YRINDVDFGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNA 357

Qy        377 PGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQ- 435
               :|||||| :|||  :||::| :|: :: :||: |:||     ||  : |:  |
Db        358 SELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPE 414

Qy        436 --RELRDWGLSFDSNLLSFSGRILQTEKI--HQGGKTFDYNPQFADWSKETRGAPLISVK 491
                : |||| :  | |:    |||:  : |  | |       |   ||| :  |   :::
Db        415 SVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTP 471

Qy        492 P--LDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRK-AIMIEVDDRTEAYLRVLQQK 548
               || | :|  :||      :|: :|::    ||:::|     |  |||| |:| :
Db        472 SDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDC 531

Qy        549 VTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQM 608
               | :| ::::||: ::  ::| :|||    |   |:| |  :|  | :::|:||||||:|:
Db        532 VRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQL 590

Qy        609 NCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASIN-EGMTRWFSRCIFQDR 667
              |||:|    | :::||  :| :| |         :|:    :||:: :  : :||
Db        591 NCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSA 650

Qy        668 GQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSI 727
               | : |   :  ||| :   :  :||||: |||||  | || |  :||   :: ||:
Db        651 FDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKT- 709

Qy        728 GRGY------NPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVS 781
               |       |:   ||| : :|||||       ||| |||::|  :| || |||::||
Db        710 --EYARVQLSPPQLAYIVVTRSMNTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVS 763

Qy        782 QAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFL 841
              | || |:|||| |||:| : |  |: :|:||||:||:|||| |  |||| |||| ||| |
Db        764 QQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATL 823

Qy        842 VGQSIHREPNLSLSNRLYYL 861
              || ::|  |  :|  : |||
Db        824 VGTNLHSIPQNALEKKFYYL 843


RESULT 6
AAU07866
ID   AAU07866 standard; Protein; 580 AA.
XX
AC   AAU07866;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Polypeptide sequence for mammalian Spg16.
XX
KW   Mammalian; reproductive-specific protein; male infertility;
KW   spermatogenesis; sperm count disorder; anti infertility; reproduction.
XX
OS   Mammalia.
XX
PN   WO200166752-A2.
XX
PD   13-SEP-2001.
XX
PF   07-MAR-2001; 2001WO-US07371.
XX
PR   07-MAR-2000; 2000US-0187518.
PR   12-JAN-2001; 2001US-0261557.
```

```
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX
PI   Wang PJ,  Page DC;
XX
DR   WPI; 2001-570774/64.
DR   N-PSDB; AAS13630.
XX
PT   Novel reproduction-specific protein, useful for treating disorders of
PT   reduced sperm count, enhancing/increasing sperm count and/or sperm
PT   activity -
XX
PS   Claim 22; Fig 16; 151pp; English.
XX
CC   The present invention relates to the isolation of novel mammalian and
CC   human reproductive-specific proteins (AAU07859-AAU07899), and the
CC   nucleic acids encoding them. The nucleic acids encoding
CC   reproductive-specific proteins are useful for diagnosing infertility
CC   which is a result of reduced sperm count, reduced sperm motility,
CC   malformed sperm or combinations of these. The sequences of the invention
CC   are useful as markers for spermatogonial cells, for identifying genes or
CC   proteins characteristic of male infertility, diagnosing or aiding in
CC   the diagnosis of infertility in men, and for contraception in which
CC   sperm production or sperm count is reduced or defective sperm is
CC   produced. Antibodies to reproductive-specific proteins are useful for
CC   determining the presence of these proteins in a sample obtained from a
CC   man being assessed for infertility, for identifying the expression of
CC   genes in particular cell type or particular developmental stage, for
CC   studies of spermatogenesis, and for immunofluorescence of germ cells or
CC   in Western blots for assessing the presence of the protein the antibody
CC   binds. The sequences of the invention are also useful for treating
CC   disorders of reduced sperm count, and for increasing sperm count and/or
CC   sperm activity. The nucleic acids of the invention are useful in gene
CC   therapy. AAU07859-AAU07882 represent the mammalian reproduction-specific
CC   proteins of the present invention.
XX
SQ   Sequence   580 AA;

  Query Match             30.1%;  Score 1365.5;  DB 22;  Length 580;
  Best Local Similarity   43.7%;  Pred. No. 4.2e-122;
  Matches  255;  Conservative  117;  Mismatches  203;  Indels    9;  Gaps    8;

Qy        283 MFNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSF 342
              |   | | :|| ||:: ||  :| :|:|:|| :|||:||:||:: || :|  :||  ::|
Db          1 MHAIYQQNKEH-FQDECSKLLVGSIVITRYNNRTYRIDDVDWNKTPKDSFVMSDGKEITF  59

Qy        343 LEYYRKQYNQEITDLKQPVLVSQP-KRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFN 401
              |||| | |   : :  ||:|: :| :|:   |  | |  :|:||| ::||: :||: ||
Db         60 LEYYSKNYGITVKEDDQPLLIHRPSERQNNHGMLLKGEILLLPELSFMTGIPEKMKKDFR 119

Qy        402 VMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEK 461
               ||||     |:|:|   :  |:  | :|:    ||  ||||   ::    ||:|  |:
Db        120 AMKDLTQQINLSPKQHHGALECLLQRISQNETASNELTRWGLSLHKDVHKIEGRLLPMER 179

Qy        462 IHQGGKTFDYNPQFADWSKE-TRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTP 520
              |:    :|    :  :| || || | :::: |:  | | | :|  : |  |:  | |:
Db        180 INLRNTSF-VTSEGLNWVKEVTRDASILTI-PMHFWALFYPKRAMDQARELVNMLEKIAG 237

Qy        521 AMGMQMRKAIMIEV-DDRTEAYLRVLQQ--KVTADTQIVVCLLSSNRKDKYDAIKKYLCT 577
               :||:      :|: ||| | |:| :|    |    |:|||::   | | | ||||  |
Db        238 PIGMRTSPPAWVELKDDRIETYIRTIQSLLGVEGKIQMVVCIIMGTRDDLYGAIKKLCCV 297

Qy        578 DCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDM 637
              | ||| :  ||:|:   : ::| || |||||:||||| |||||| :|::|:| |||
Db        298 QSPVPSQVINVRTIGQPTRLRSVAQKILLQMNCKLGGELWGVDIPLKQLMVIGMDVYHDP 357

Qy        638 TAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRI 697
              : | ||: ||||||| :|:|:|| :||   ||:|| ||:|| :|: :   |  :| :|
Db        358 SRGMRSVVGFVASINLTLTKWYSRVVFQMPHQEIVDSLKLCLVGSLKKYYEVNHCLPEKI 417

Qy        698 IVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQ 757
              :|||||| |||||: |||:||   | ::    |:|:  | ||:|:::|  :  :
Db        418 VVYRDGVSDGQLKTVANYEIPQLQKCFEAFD-NYHPKMVVFVVQKKISTNLYLAAPDHFV 476

Qy        758 NPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCH 817
              | ||||:|  :|  || ||:::: || |  ||||  : : : | |||:||||:||||
Db        477 TPSPGTVVDHTITSCEWVDFYLLAHHVRQGCGIPTHYICVLNTANLSPDHMQRLTFKLCH 536

Qy        818 IYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 861
              :|:|||| |||||||:|||||||| || :| || : |   |::|
Db        537 MYWNWPGTIRVPAPCKYAHKLAFLSGQILHHEPAIQLCGNLFFL 580


RESULT 7
AAB94209
ID   AAB94209 standard; Protein; 530 AA.
XX
AC   AAB94209;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:14559.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 14559; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
```

```
SQ   Sequence   530 AA;

  Query Match             26.9%;  Score 1220;  DB 22;  Length 530;
  Best Local Similarity   43.5%;  Pred. No. 3.9e-108;
  Matches  231;  Conservative   98;  Mismatches  196;  Indels    6;  Gaps    5;

Qy        335 ADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRR-GPGGTLPGPAMLIPELCYLTGLT 393
              :||  ::|||||| | |   : :  ||:|: :|  |:    |  | |  :|:||| ::||:
Db          2 SDGKEITFLEYYSKNYGITVKEEDQPLLIHRPSERQDNHGMLLKGEILLLPELSFMTGIP 61

Qy        394 DKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFS 453
              :||: ||  |||||    |:|:|    :  |:  | ||:    ||  |||    ::
Db         62 EKMKKDFRAMKDLAQQINLSPKQHHSALECLLQRIAKNEAATNELMRWGLRLQKDVHKIE 121

Qy        454 GRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQ 513
              ||:|  |:|:    :|     |  :| ||    | |   |:  | | | :|  : |  |:
Db        122 GRVLPMERINLKNTSF-ITSQELNWVKEVTRDPSILTIPMHFWALFYPKRAMDQARELVN 180

Qy        514 NLFKVTPAMGMQMRKAIMIEV-DDRTEAYLRVLQQKVTAD--TQIVVCLLSSNRKDKYDA 570
                | |:   :||:|     :|: ||| | |:| :|  : |:   |:|||::   | | | |
Db        181 MLEKIAGPIGMRMSPPAWVELKDDRIETYVRTIQSTLGAEGKIQMVVCIIMGPRDDLYGA 240

Qy        571 IKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVG 630
              |||  |   | ||| |  ||:|:   : ::| || ||:|||:|||||| |||||| :|::|
Db        241 IKKLCCVQSPVPSQVVNVRTIGQPTRLRSVAQKILLQINCKLGGELWGVDIPLKQLMVIG 300

Qy        631 IDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCN 690
              :| ||| : | ||: |||||||  :|:|:|| :||   ||:|| ||:||  :|: :   |
Db        301 MDVYHDPSRGMRSVVGFVASINLTLTKWYSRVVFQMPHQEIVDSLKLCLVGSLKKFYEVN 360

Qy        691 EYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFA 750
                :| :|:|||||| |||||: |||:||   | ::    | |:  | ||:|:::|  :
Db        361 HCLPEKIVVYRDGVSDGQLKTVANYEIPQLQKCFEAF-ENYQPKMVVFVVQKKISTNLYL 419

Qy        751 QSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQR 810
              :      | ||||:|  :|  || ||::::  || |   ||||  : : : | |||:||
Db        420 AAPQNFVTPTPGTVVDHTITSCEWVDFYLLAHHVRQGCGIPTHYVCVLNTANLSPDHMQR 479

Qy        811 LTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 861
              ||:||||:|:|||| |||||||:|||||||| |  :|  | : |   |::|
Db        480 LTFKLCHMYWNWPGTIRVPAPCKYAHKLAFLSGHILHHGPAIQLCENLFFL 530


RESULT 8
AAB92722
ID   AAB92722 standard; Protein; 371 AA.
XX
AC   AAB92722;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:11144.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 11144; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   371 AA;

  Query Match             21.2%;  Score 959;  DB 22;  Length 371;
  Best Local Similarity   47.2%;  Pred. No. 3.1e-83;
  Matches  174;  Conservative   69;  Mismatches  122;  Indels    4;  Gaps    3;

Qy        496 WLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEV-DDRTEAYLRVLQQKVTAD-- 552
              | | | :|  : |  |:  | |:   :||:|     :|: ||| | |:| :|  : |:
Db          4 WALFYPKRAMDQARELVNMLEKIAGPIGMRMSPPAWVELKDDRIETYVRTIQSTLGAEGK 63

Qy        553 TQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKM 612
               |:|||::   | | | || |  |   | ||| |  ||:|:   : ::| || ||:|||:
Db         64 IQMVVCIIMGPRDDLYGAINKLCCVQSPVPSQVVNVRTIGQPTRLRSVAQKILLQINCKL 123

Qy        613 GGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELV 672
              ||||| |||||| :|:::|:| ||| : | ||: |||||||  :|:|:|| :||   ||:|
Db        124 GGELWGVDIPLKQLMVIGMDVYHDPSRGMRSVVGFVASINLTLTKWYSRVVFQMPHQEIV 183

Qy        673 DGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYN 732
              | ||:||  :|: :   |  :| :|:||||||  ||||||: |||:||   | ::    |
Db        184 DSLKLCLVGSLKKFYEVNHCLPEKIVVYRDGVSDGQLKTVANYEIPQLQKCFEAF-ENYQ 242

Qy        733 PRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPT 792
              |:  | ||:|:::|  :  :      | ||||:|  :|  || ||::::  || |   ||
Db        243 PKMVVFVVQKKISTNLYLAAPQNFVTPTPGTVVDHTITSCEWVDFYLLAHHVRQGCGIPT 302

Qy        793 HYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNL 852
              ||  : : : | |||:||||:||||:|:|||| |||||||:|||||||| |  :| || :
Db        303 HYVCVLNTANLSPDHMQRLTFKLCHMYWNWPGTIRVPAPCKYAHKLAFLSGHILHHEPAI 362

Qy        853 SLSNRLYYL 861
               |   |::|
Db        363 QLCENLFFL 371


RESULT 9
AAB58977
ID   AAB58977 standard; Protein; 130 AA.
XX
AC   AAB58977;
```

```
XX
DT   27-MAR-2001  (first entry)
XX
DE   Breast and ovarian cancer associated antigen protein sequence SEQ ID 685.
XX
KW   Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW   nootropic; neurpprotective; antiviral; antiallergic; hepatotropic;
KW   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW   antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW   Addison's disease; allergy; autoimmune haemolytic anaemia;
KW   autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW   multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW   cardiovascular disorder; wound healing; neurological disease.
XX
OS   Homo sapiens.
XX
PN   WO200055173-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05881.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-611515/58.
DR   N-PSDB; AAF21880.
XX
PT   New human breast and ovarian cancer associated gene sequences and the
PT   polypeptides encoded by these genes, useful in the prevention,
PT   treatment and diagnosis of cancer, immune disorders, cardiovascular
PT   disorders and neurological diseases -
XX
PS   Claim 11; Page 1139; 1299pp; English.
XX
CC   Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC   proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC   associated with breast and ovarian cancer. Included in the invention are
CC   sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC   isolation and characterisation of the DNA and protein sequences of the
CC   invention. The breast and ovarian cancer associated DNA, protein, agonist
CC   or antagonist sequences exhibit cytostatic; immunosuppressive;
CC   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
CC   antibacterial; antifungal; antiparasitic and cardiant activity. The
CC   polynucleotide and protein sequences are used in the diagnosis of cancer,
CC   particularly breast and ovarian cancer. The nucleic acid sequences,
CC   proteins, agonists and agonists may also be used in the diagnosis,
CC   prevention and treatment of immune disorders e.g. Addison's disease,
CC   allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC   diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC   arthritis and ulcerative colitis; cardiovascular disorders such as
CC   myocardial ischaemias; wound healing; neurological diseases such as
CC   cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ   Sequence   130 AA;

  Query Match             14.4%;  Score 652;  DB 21;  Length 130;
  Best Local Similarity   99.2%;  Pred. No. 2.1e-54;
  Matches  123;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        578 DCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDM 637
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db          5 DCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCXHDM 64

Qy        638 TAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRI 697
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 TAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRI 124

Qy        698 IVYR 701
                ||||
Db        125 IVYR 128


RESULT 10
ABG05957
ID   ABG05957 standard; Protein; 341 AA.
XX
AC   ABG05957;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #5948.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS70144.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 36316; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   341 AA;

  Query Match             12.3%;  Score 559;  DB 22;  Length 341;
  Best Local Similarity   42.4%;  Pred. No. 9.8e-45;
  Matches  111;  Conservative   50;  Mismatches   67;  Indels   34;  Gaps    3;

Qy        175 QQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYN 234
              |  |||: |:|: || :::|||| |||  :|| |:| :|||||
Db         16 QPTVTELSSETQRGETIKMTITLKRELPSSSPVCIQVFNIIFR---------------- 58
```

```
Qy        235 PNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHK 294
                      :| :|||| |:  :|  ::   |||:||||:||||:||     :|
Db         59 ----------KLSLWPGFAISVSYFERKLLFSADVSYKVLRNETVLEFMTALCQRTGLSC 108

Qy        295 FQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEI 354
              | :   |: |||:|||:|| :|| :|||||   |  ||:| ||: :::::||::||:  :
Db        109 FTQTCEKQLIGLIVLTRYNNRTYSIDDIDWSVKPTHTFQKRDGTEITYVDYYKQQYDITV 168

Qy        355 TDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTP 414
              :|| ||:|||  |::|         | ||||||:||||||:  :|| :|| :|  |||:|
Db        169 SDLNQPMLVSLLKKKRNDNSEAQ-LAHLIPELCFLTGLTDQATSDFQLMKAVAEKTRLSP 227

Qy        415 EQRQREVGRLIDYIHKNDNVQR 436
                ||: : ||:      ||:||
Db        228 SGRQQRLARLV------DNIQR 243
```

RESULT 11
AAG51562
ID AAG51562 standard; Protein; 983 AA.
XX
AC AAG51562;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 65455.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.

PR 29-OCT-1999; 99US-0162142.

Query Match 11.5%; Score 521; DB 21; Length 983;
Best Local Similarity 23.2%; Pred. No. 2.7e-40;
Matches 220; Conservative 150; Mismatches 424; Indels 154; Gaps 38;

```
Qy          4 RARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRGTAGGTAKSQ 63
              |:| | ||   |::: |        :|   :||  |||::               | |
Db         60 RSRRRNRG---GRKSDQ---GDVCMRPS-SRPRKPPPPSQ--------------TTSSA 97

Qy         64 GLQISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTS 123
                :|| |:            | : :  |:| :       |: |    :  ||| |
Db         98 VSVATAG--EIVAVN--------HQMQMGVRKNSNFAPRPGFGTLGTKCIVKANHF-LAD 146

Qy        124 RPQWALYQYHIDYNPLMEARRLRSAL------LFQHEDLIGKCHAFD--------GTILF 169
               |   | || :   | : :: :  |:      |::  ||  :  |:|        | : |
Db        147 LPTKDLNQYDVTITPEVSSKSVNRAIIAELVRLYKESDLGRRLPAYDGRKSLYTAGELPF 206

Qy        170 LPKRLQQKVTE----VFSKTRNGEDVRITIT------------LTNELPPTSPTCLQFY 212
               |    |: :    : :  :     :: |              |  :        :|
Db        207 TWKEFSVKIVDEDDGIINGPKRERSYKVAIKFVARANMHHLGEFLAGKRADCPQEAVQIL 266

Qy        213 NIIFRRLLKIMNLQQIGRNYYNPNDPIDIP---SHRLVIWPGFTTSILQYENSIMLCTDV 269
              :|: |  | :     :||::::|:  |  |       |  | ||  ||   :  : |  |:
Db        267 DIVLRE-LSVKRFCPVGRSFFSPD--IKTPQRLGEGLESWCGFYQSIRPTQMGLSLNIDM 323

Qy        270 -SHKVLRSETVLDFMFNFYHQ-------TEEHKFQEQVSKEXIGLVVLTKYNI-KTYRVD 320
               |   :   |::|:     :        ::  : : :     : : |  : |: : |||
Db        324 ASAAFIEPLPVIEFVAQLLGKDVLSKPLSDSDRVKIKKGLRGVKVEVTHRANVRRKYRVA 383

Qy        321 DIDWDQNPKSTFKKADGSGV-SFLEYYRKQYNQEITDLKQPVL-VSQPKRRRGPGGTLPG 378
              :       :  |   :   : | :||::: |   |      | | |   |:       ||
Db        384 GLTTQPTRELMFPVDENCTMKSVIEYFQEMYGFTIQHTHLPCLQVGNQKK----ASYLPM 439

Qy        379 PAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQREL 438
               |  | |    |    |  |:  :   | | |   |  |: :: | :   |   :
Db        440 EACKIVEGQRYT----KRLNEKQITALLKV-TCQRPRDRENDILRTVQ--HNAYDQDPYA 492

Qy        439 RDWGLSFDSNLLSFSGRILQTE--KIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNW 496
              :::|::    | |   |||     | |: || |  ||   |:   :   :|:   :  |
Db        493 KEFGMNISEKLASVEARILPAPWLKYHENGKEKDCLPQVGQWNMMNK--KMINGMTVSRW 550

Qy        497 LLIYTRRNYE--AANSLIQNLFKVTPAMGMQMRKAIMIEV----DDRTE-----AYLRVL 545
              :   |: :   |      | ::    ||:     :| :     |: |       |   :
Db        551 ACVNFSRSVQENVARGFCNELGQMCEVSGMEFNPEPVIPIYSARPDQVEKALKHVYHTSM 610

Qy        546 QQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVAR---TLGKQQTVMAIAT 602
               :    : :::: :|  |    |  :|:  |:    ||| : :    : ||
Db        611 NKTKGKELELLLAILPDNNGSLYGDLKRICETELGLISQCCLTKHVFKISKQ-----YLA 665

Qy        603 KIALQMNCKMGGE--------LWRV----DIPLKLVMIVGIDCYHDMTA--GRRSIAGFV 648
              ::|::| ||||           |:    |||    :| | |  |          |||  |
Db        666 NVSLKINVKMGGRNTVLVDAISCRIPLVSDIP---TIIFGADVTHPENGEESSPSIAAVV 722

Qy        649 ASIN-EGMTRWFSRCIFQDRGQELVDGLKVCLQAALR-------------AWNSCNEYMP 694
              || :   :|::      |   |||:  |    |  :|                ::      |
Db        723 ASQDWPEVTKYAGLVCAQAHRQELIQDLYKTWQDPVRGTVSGGMIRDLLISFRKATGQKP 782

Qy        695 SRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFA---- 750
               ||| ||||| :||   :: ||:        |:   | |  | |||:|| :|| ||
Db        783 LRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEPNYQPPVTFIVVQKRHHTRLFANNHR 842

Qy        751 --QSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHI 808
                |  |  | |||||:| ::  |  :||:: | |   |:  | ||:|::|  :    | |
Db        843 DKNSTDRSGNILPGTVVDTKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDENNFTADGI 902

Qy        809 QRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSN 856
              | ||  ||: |       : :  |  |||  ||      : || :   |
Db        903 QSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRA--RFYLEPEIMQDN 948
```

RESULT 12
AAG51561

ID AAG51561 standard; Protein; 988 AA.
XX
AC AAG51561;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 65454.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.

```
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match           11.5%;  Score 521;  DB 21;  Length 988;
  Best Local Similarity  23.2%;  Pred. No. 2.7e-40;
  Matches  220;  Conservative  150;  Mismatches  424;  Indels  154;  Gaps  38;

Qy          4 RARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRGTAGGTAKSQ 63
              |:| | ||   |::: |       :|   :|| |||::               |  |
Db         65 RSRRRNRG---GRKSDQ---GDVCMRPS-SRPRKPPPPSQ--------------TTSSA 102

Qy         64 GLQISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTS 123
                :||  |:            | : :  |:| :       |: |    :  ||| |
Db        103 VSVATAG--EIVAVN--------HQMQMGVRKNSNFAPRPGFGTLGTKCIVKANHF-LAD 151

Qy        124 RPQWALYQYHIDYNPLMEARRLRSAL------LFQHEDLIGKCHAFD--------GTILF 169
                |   | || :   | : :: :  |:      |::  ||  :  |:|        | : |
Db        152 LPTKDLNQYDVTITPEVSSKSVNRAIIAELVRLYKESDLGRRLPAYDGRKSLYTAGELPF 211

Qy        170 LPKRLQQKVTE----VFSKTRNGEDVRITIT-------------LTNELPPTSPTCLQFY 212
                |    |: :    : :  :     :: |             |  :        :|

Db        212 TWKEFSVKIVDEDDGIINGPKRERSYKVAIKFVARANMHHLGEFLAGKRADCPQEAVQIL 271

Qy        213 NIIFRRLLKIMNLQQIGRNYYNPNDPIDIP---SHRLVIWPGFTTSILQYENSIMLCTDV 269
              :|: |  | :     :||::::|:  |  |      |  | ||  ||   :  : |  |:
Db        272 DIVLRE-LSVKRFCPVGRSFFSPD--IKTPQRLGEGLESWCGFYQSIRPTQMGLSLNIDM 328

Qy        270 -SHKVLRSETVLDFMFNFYHQ-------TEEHKFQEQVSKEXIGLVVLTKYNI-KTYRVD 320
               |   :    |::|:     :       ::  : : :     : | : |: : |||
Db        329 ASAAFIEPLPVIEFVAQLLGKDVLSKPLSDSDRVKIKKGLRGVKVEVTHRANVRRKYRVA 388

Qy        321 DIDWDQNPKSTFKKADGSGV-SFLEYYRKQYNQEITDLKQPVL-VSQPKRRRGPGGTLPG 378
              :       :  |   :   : | :||::: |   |      | | |   |:       ||
Db        389 GLTTQPTRELMFPVDENCTMKSVIEYFQEMYGFTIQHTHLPCLQVGNQKK----ASYLPM 444

Qy        379 PAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQREL 438
               |  | |     |    |  |:  :   | | |   |  |: :: | :   |   :
Db        445 EACKIVEGQRYT----KRLNEKQITALLKV-TCQRPRDRENDILRTVQ--HNAYDQDPYA 497

Qy        439 RDWGLSFDSNLLSFSGRILQTE--KIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNW 496
              :::|::    | |   |||     | |: || |  ||   |:   :   :|:   :  |
Db        498 KEFGMNISEKLASVEARILPAPWLKYHENGKEKDCLPQVGQWNMMNK--KMINGMTVSRW 555

Qy        497 LLIYTRRNYE--AANSLIQNLFKVTPAMGMQMRKAIMIEV----DDRTE-----AYLRVL 545
              :   |: :   |      | ::    ||:     :| :     |: |      |  :
Db        556 ACVNFSRSVQENVARGFCNELGQMCEVSGMEFNPEPVIPIYSARPDQVEKALKHVYHTSM 615

Qy        546 QQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVAR---TLGKQQTVMAIAT 602
              :    : :::: :|  |    |  :|:   |:    ||| : :    : ||
Db        616 NKTKGKELELLLAILPDNNGSLYGDLKRICETELGLISQCCLTKHVFKISKQ-----YLA 670

Qy        603 KIALQMNCKMGGE--------LWRV----DIPLKLVMIVGIDCYHDMTA--GRRSIAGFV 648
              ::|::| ||||          |:   |||    :| | |  |          |||  |
Db        671 NVSLKINVKMGGGRNTVLVDAISCRIPLVSDIP---TIIFGADVTHPENGEESSPSIAAVV 727

Qy        649 ASIN-EGMTRWFSRCIFQDRGQELVDGLKVCLQAALR------------AWNSCNEYMP 694
              || :   :|::      |   |||:  |    |  :|             ::      |
Db        728 ASQDWPEVTKYAGLVCAQAHRQELIQDLYKTWQDPVRGTVSGGMIRDLLISFRKATGQKP 787

Qy        695 SRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFA---- 750
              ||| ||||| :||   :: ||:        |:   | |  | |||:|| :|| ||
Db        788 LRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEPNYQPPVTFIVVQKRHHTRLFANNHR 847

Qy        751 --QSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHI 808
                |  | | |||||:| ::  |  :||:: | |   |:  | ||:|::| :    | |
Db        848 DKNSTDRSGNILPGTVVDTKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDENNFTADGI 907

Qy        809 QRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSN 856
              | ||  ||: |      : :  |  |||  ||      : || :   |
Db        908 QSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRA--RFYLEPEIMQDN 953


RESULT 13
AAG42118
ID   AAG42118 standard; Protein; 1048 AA.
XX
AC   AAG42118;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 52487.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
```

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

```
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             11.4%;  Score 517;  DB 21;  Length 1048;
  Best Local Similarity   23.4%;  Pred. No. 7.2e-40;
  Matches  238;  Conservative  159;  Mismatches  428;  Indels  190;  Gaps   40;

Qy          1 MTGRARARARG---RARGQETAQLVGSTASQQP--------GYIQPRPQP---------- 39
              ::|  |   ||   :  ||      |   : ||         || ||  |
Db         25 VSGGGRGSQRGGFQQGGGQHQG---GRGYTPQPQQGRGGRGYGQPPQQQQQYGGPQEYQ 81

Qy         40 ------PPAEGEL--FGRGRQRGTAGGTAKSQGLQ------------------------ 66
                    || :|    :| ||  | : | .: | :
Db         82 GRGRGGPPHQGGRGGYGGGRGGGPSSGPPQRQSVPELHQATSPTYQAVSSQPTLSEVSPT 141

Qy         67 -------ISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHF 119
                     ::  |::||  |:|   :   :  :::     ::  | | ||    :  |||
Db        142 QVPEPTVLAQQFEQLS-VEQGAPSQAIQPIPSSSKAFKFPMRPGK-GQSGKRCIVKANHF 199

Qy        120 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCH------AFDG-----TIL 168
                  |   |: | :   | : :| :  |:: |  |      |      |:||     |
Db        200 -FAELPDKDLHHYDVTITPEVTSRGVNRAVMKQLVDNYRDSHLGSRLPAYDGRKSLYTAG 258

Qy        169 FLP------------------KRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCL 209
               ||                  :| :::   |       :   : : |  :        |
Db        259 PLPFNSKEFRINLLDEEVGAGGQRREREFKVVIKLVARADLHHLGMFLEGKQSDAPQEAL 318

Qy        210 QFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHR-----LVIWPGFTTSILQYENSIM 264
              |  :|: |  |        :||::|:|    ||   :     |  | ||  ||   :  :
Db        319 QVLDIVLRE-LPTSRYIPVGRSFYSP----DIGKKQSLGDGLESWRGFYQSIRPTQMGLS 373

Qy        265 LCTDVSHKV-LRSETVLDFMFNFYHQ-------TEEHKFQEQVSKEXIGLVVLTKYNI-K 315
              |  |:|   : :  |: |: :  ::       ::  : : : :   : : |  : |: :
Db        374 LNIDMSSTAFIEANPVIQFVCDLLNRDISSRPLSDADRVKIKKALRGVKVEVTHRGNMRR 433

Qy        316 TYRVDDIDWDQNPKSTFKKAD-GSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGG 374
               ||:  :      : ||   :  :  | :||: : |   |   : | |      |
Db        434 KYRISGLTAVATRELTFPVDERNTQKSVVEYFHETYGFRIQHTQLPCLQVGNSNR----- 488

Qy        375 TLPGPAMLIPELCYLT--GLTDKMRNDFNVMKDLAVHTRLTPEQRQREV---GRLIDYIH 429
                  |  |  |:| :       |  |:  :   | | |   |  |::::    :| ||

Db        489 ----PNYLPMEVCKIVEGQRYSKRLNERQITALLKV-TCQRPIDREKDILQTVQLNDY-- 541

Qy        430 KNDNVQRELRDWGLSFDSNLLSFSGRILQTE--KIHQGGKTFDYNPQFADWSKETRGAPL 487
              ||  :|   :|:   ::| |   |||     | |: |:      ||   :   :
Db        542 AKDNYAQE---FGIKISTSLASVEARILPPPWLKYHESGREGTCLPQVGQWNMMNK--KM 596

Qy        488 ISVKPLDNWLLIYTRRNYE--AANSLIQNLFKVTPAMGMQMR-KAIMIEVDDRTEAYLRV 544
              |:   ::||: |   |  :   | :  | | ::    ||    : ::  |  | |   :|
Db        597 INGGTVNNWICINFSRQVQDNLARTFCQELAQMCYVSGMAFNPEPVLPPVSARPEQVEKV 656

Qy        545 LQQKVTADT---------QIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGK-Q 594
              |: :    |           ::: :|  |    |  :|:   |:    ||| : : : |
Db        657 LKTRYHDATSKLSQGKEIDLLIVILPDNNGSLYGDLKRICETELGIVSQCCLTKHVFKMS 716

Qy        595 QTVMAIATKIALQMNCKMGG-ELWRVD-----IPL---KLVMIVGIDCYHDMTA--GRRS 643
              :  ||    :||::| |:||     ||     |||   :  :| | |  |         |
Db        717 KQYMA---NVALKINVKVGGRNTVLVDALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPS 773

Qy        644 IAGFVASIN-EGMTRWFSRCIFQDRGQELVDGL-------------KVCLQAALRAWNSC 689
              ||  |||  :   :|::       |   |||:  |           ::  | |:
Db        774 IAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFKEWKDPQKGVVTGGMIKELLIAFRRS 833

Qy        690 NEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFF 749
              : | ||| |||||  :||   :: ||:         |:  || |  | :||:|| :|| |
Db        834 TGHKPLRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEAGYQPPVTFVVVQKRHHTRLF 893

Qy        750 AQ------SGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGL 803
              ||      |  |  | |||||:| ::   |  :||:: | |    |:  | ||:|::| :
Db        894 AQNHNDRHSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDENNF 953

Qy        804 KPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRL 858
              | :| ||  ||: |     : :  |  |||  ||      : ||  | |  :
Db        954 TADGLQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRA--RFYMEPETSDSGSM 1006


RESULT 14
AAG51563
ID   AAG51563 standard; Protein; 906 AA.
XX
AC   AAG51563;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 65456.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
```

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

```
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             11.3%;  Score 514.5;  DB 21;  Length 906;
  Best Local Similarity   23.1%;  Pred. No. 9.9e-40;
  Matches  212;  Conservative  145;  Mismatches  414;  Indels  147;  Gaps   35;

Qy         34 QPRPQPPPAEGELFGRGRQRGTAGGTAKSQGLQISAGFQELSXAERGGRRRDFHDLGVNT 93
              :||  |||::           |  |     :||  |:          | : :
Db          6 RPRKPPPPSQ--------------TTSSAVSVATAG--EIVAVN--------HQMQMGV 40

Qy         94 RQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSAL---- 149
              |:| :       |: |    :  ||| |   |   | || :   | : :: :  |:
Db         41 RKNSNFAPRPGFGTLGTKCIVKANHF-LADLPTKDLNQYDVTITPEVSSKSVNRAIIAEL 99

Qy        150 --LFQHEDLIGKCHAFD--------GTILFLPKRLQQKVTE----VFSKTRNGEDVRITI 195
               |::  ||  :  |:|        | : |  |    |: :    : :  :     :: |
Db        100 VRLYKESDLGRRLPAYDGRKSLYTAGELPFTWKEFSVKIVDEDDGIINGPKRERSYKVAI 159

Qy        196 T-------------LTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIP 242
                          |  :       :|  :|: |  | :     :||::::|:  |  |
Db        160 KFVARANMHHLGEFLAGKRADCPQEAVQILDIVLRE-LSVKRFCPVGRSFFSPD--IKTP 216

Qy        243 ---SHRLVIWPGFTTSILQYENSIMLCTDV-SHKVLRSETVLDFMFNFYHQ-------TE 291
                 |  | ||  ||   :  : |  |: |   :    |::|:      :        ::
Db        217 QRLGEGLESWCGFYQSIRPTQMGLSLNIDMASAAFIEPLPVIEFVAQLLGKDVLSKPLSD 276

Qy        292 EHKFQEQVSKEXIGLVVLTKYNI-KTYRVDDIDWDQNPKSTFKKADGSGV-SFLEYYRKQ 349
              : : :     : : |  : |: : |||  :       :  |   :   : | :||:::
Db        277 SDRVKIKKGLRGVKVEVTHRANVRRKYRVAGLTTQPTRELMFPVDENCTMKSVIEYFQEM 336

Qy        350 YNQEITDLKQPVL-VSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAV 408
              |   |      | | |   |:        ||  |  | |     |     |  |:  :   | |
Db        337 YGFTIQHTHLPCLQVGNQKK----ASYLPMEACKIVEGQRYT----KRLNEKQITALLKV 388

Qy        409 HTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTE--KIHQGG 466
               |   |  |: :: | :   |   :     :::|::    | |   |||     | |: |
Db        389 -TCQRPRDRENDILRTVQ--HNAYDQDPYAKEFGMNISEKLASVEARILPAPWLKYHENG 445

Qy        467 KTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYE--AANSLIQNLFKVTPAMGM 524
              |  |  ||   |:   :   :|:   :  |  :   |: :   |       | ::    ||
Db        446 KEKDCLPQVGQWNMMNK--KMINGMTVSRWACVNFSRSVQENVARGFCNELGQMCEVSGM 503

Qy        525 QMRKAIMIEV----DDRTE-----AYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYL 575
              :    :| :     |: |      |   : :    : :::: :|  |     |  :|:
Db        504 EFNPEPVIPIYSARPDQVEKALKHVYHTSMNKTKGKELELLLAILPDNNGSLYGDLKRIC 563

Qy        576 CTDCPTPSQCVVAR---TLGKQQTVMAIATKIALQMNCKMGGE--------LWRV----D 620
               |:    ||| : :    : ||        ::|::| ||||           |:    |
Db        564 ETELGLISQCCLTKHVFKISKQ-----YLANVSLKINVKMGGRNTVLVDAISCRIPLVSD 618

Qy        621 IPLKLVMIVGIDCYHDMTA--GRRSIAGFVASIN-EGMTRWFSRCIFQDRGQELVDGLKV 677
              ||    :| | |  |          |||  ||| :   :|::      |   |||:  |
Db        619 IP---TIIFGADVTHPENGEESSPSIAAVVASQDWPEVTKYAGLVCAQAHRQELIQDLYK 675

Qy        678 CLQAALR-------------AWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCL 724
               |  :|              ::       | ||| |||||  :||   :: ||:
```

```
Db        676 TWQDPVRGTVSGGMIRDLLISFRKATGQKPLRIIFYRDGVSEGQFYQVLLYELDAIRKAC 735

Qy        725 KSIGRGYNPRXTVIVVKKRVNTRFFA------QSGGRLQNPLPGTVIDVEVTRPEWYDFF 778
               |:   | |  | |||:|| :|| ||       |  |  | |||||:| ::  |  :||:
Db        736 ASLEPNYQPPVTFIVVQKRHHTRLFANNHRDKNSTDRSGNILPGTVVDTKICHPTEFDFY 795

Qy        779 IVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKL 838
               : | |    |:  | ||:|::|  : :  | || ||  ||: |      : :  |  |||
Db        796 LCSHAGIQGTSRPAHYHVLWDENNFTADGIQSLTNNLCYTYARCTRSVSIVPPAYYAHLA 855

Qy        839 AFLVGQSIHREPNLSLSN 856
               ||      : || :   |
Db        856 AFRA--RFYLEPEIMQDN 871


RESULT 15
AAG52593
ID   AAG52593 standard; Protein; 1064 AA.
XX
AC   AAG52593;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 66875.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
```

PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.2%; Score 510; DB 21; Length 1064;

Best Local Similarity 23.3%; Pred. No. 3.5e-39;
Matches 240; Conservative 161; Mismatches 425; Indels 204; Gaps 41;

```
Qy          1 MTGRARARARG---RARGQETAQLVGSTASQQP--------GYIQPRPQP---------- 39
              ::|  |   ||   :  ||      |   : ||        || ||  |
Db         25 VSGGGRGSQRGGFQQGGGQHQG---GRGYTPQPQQGGRGGRGYGQPPQQQQQYGGPQEYQ 81

Qy         40 ------PPAEGEL--FGRGRQRGTAGGTAKSQGLQ------------------------- 66
                  || :|    :| ||  | : |  : | :
Db         82 GRGRGGPPHQGGRGGYGGGRGGGPSSGPPQRQSVPELHQATSPTYQAVSSQPTLSEVSPT 141

Qy         67 -------ISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHF 119
                    ::  |::|| |:|   :    :  :::     ::  | | ||    :  |||
Db        142 QVPEPTVLAQQFEQLS-VEQGAPSQAIQPIPSSSKAFKFPMRPGK-GQSGKRCIVKANHF 199

Qy        120 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCH------AFDG-----TIL 168
                |   |: | :   | : :| :  |:: |  |       |      |:||     |
Db        200 -FAELPDKDLHHYDVTITPEVTSRGVNRAVMKQLVDNYRDSHLGSRLPAYDGRKSLYTAG 258

Qy        169 FLP------------------KRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCL 209
              ||                 :| :::   |       :   : : |  :        |
Db        259 PLPFNSKEFRINLLDEEVGAGGQRREREFKVVIKLVARADLHHLGMFLEGKQSDAPQEAL 318

Qy        210 QFYNIIFRRL------LKIMNL---------QQIGRNYYNPNDPIDIPSHR-----LVIW 249
              |  :|: | |      | |: :           :||::|:|    ||   :     |  |
Db        319 QVLDIVLRELPTSRYWLLILWMFASLLPCLYIPVGRSFYSP----DIGKKQSLGDGLESW 374

Qy        250 PGFTTSILQYENSIMLCTDVSHKV-LRSETVLDFMFNFYHQ-------TEEHKFQEQVSK 301
              '||  ||   :  : |  |:|    : :  |: |: :  ::       ::  : : : :
Db        375 RGFYQSIRPTQMGLSLNIDMSSTAFIEANPVIQFVCDLLNRDISSRPLSDADRVKIKKAL 434

Qy        302 EXIGLVVLTKYNI-KTYRVDDIDWDQNPKSTFKKAD-GSGVSFLEYYRKQYNQEITDLKQ 359
              : : |  : |: : ||:  :       : ||   :  :  | :||: : |   |   :
Db        435 RGVKVEVTHRGNMRRKYRISGLTAVATRELTFPVDERNTQKSVVEYFHETYGFRIQHTQL 494

Qy        360 PVLVSQPKRRRGPGGTLPGPAMLIPELCYLT--GLTDKMRNDFNVMKDLAVHTRLTPEQR 417
              | |      |       |  |  |:| :       |  |:  :   | | |   |  |
Db        495 PCLQVGNSNR---------PNYLPMEVCKIVEGQRYSKRLNERQITALLKV-TCQRPIDR 544

Qy        418 QREV---GRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTE--KIHQGGKTFDYN 472
              ::::    :| ||    ||  :|   :|:   ::| |   |||     | |: |:
Db        545 EKDILQTVQLNDY--AKDNYAQE---FGIKISTSLASVEARILPPPWLKYHESGREGTCL 599

Qy        473 PQFADWSKETRGAPLISVKPLDNWLLIYTRRNYE--AANSLIQNLFKVTPAMGMQMR-KA 529
              ||   |:   :  :|:   ::||: |   |  :   | :  | | ::    ||    :
Db        600 PQVGQWNMMNK--KMINGGTVNNWICINFSRQVQDNLARTFCQELAQMCYVSGMAFNPEP 657

Qy        530 IMIEVDDRTEAYLRVLQQKVTADT---------QIVVCLLSSNRKDKYDAIKKYLCTDCP 580
              ::  |  | |   :||: :    |           ::: :|  |    |  :|:   |:
Db        658 VLPPVSARPEQVEKVLKTRYHDATSKLSQGKEIDLLIVILPDNNGSLYGDLKRICETELG 717

Qy        581 TPSQCVVARTLGK-QQTVMAIATKIALQMNCKMGG-ELWRVD-----IPL---KLVMIVG 630
              |||  :  :  |  :  ||    :||::| |:||     ||     |||   :  :| |
Db        718 IVSQCCLTKHVFKMSKQYMA---NVALKINVKVGGRNTVLVDALSRRIPLVSDRPTIIFG 774

Qy        631 IDCYHDMTA--GRRSIAGFVASIN-EGMTRWFSRCIFQDRGQELVDGL------------ 675
              |  |          |||  ||| :   :|::      |   |||:  |
Db        775 ADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFKEWKDPQKGVV 834

Qy        676 -KVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPR 734
               ::  | |:     : | ||| |||||  :||   :: ||:          |:  || |
Db        835 TGGMIKELLIAFRRSTGHKPLRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEAGYQPP 894

Qy        735 XTVIVVKKRVNTRFFAQ------SGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGS 788
              | :||:|| :|| |||       |  |  | |||||:| ::  |  :||:: | |   |:
Db        895 VTFVVVQKRHHTRLFAQNHNDRHSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGT 954

Qy        789 VSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHR 848
              | ||:|::| :    | :| ||  ||: |      : :  |  |||  ||      :
Db        955 SRPAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRA--RFYM 1012

Qy        849 EPNLSLSNRL 858
              ||  | |  :
```

Db 1013 EPETSDSGSM 1022

Search completed: July 22, 2003, 07:01:05
Job time : 32.1988 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 06:59:33 ; Search time 11.4084 Seconds
(without alignments)
2220.565 Million cell updates/sec

Title: US-09-873-737A-6
Perfect score: 4534
Sequence: 1 MTGRARARARGRARGQETAQ.....VGQSIHREPNLSLSNRLYYL 861

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :. Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	112.5	2.5	3457	2	US-08-416-603-4	Sequence 4, Appli
2	112	2.5	1180	3	US-09-224-024-28	Sequence 28, Appl
3	112	2.5	1180	5	PCT-US94-07902-28	Sequence 28, Appl
4	110.5	2.4	560	2	US-08-756-317-11	Sequence 11, Appl
5	110	2.4	1253	1	US-08-252-966B-12	Sequence 12, Appl
6	110	2.4	1261	1	US-08-252-966B-18	Sequence 18, Appl
7	107.5	2.4	615	4	US-09-134-001C-3284	Sequence 3284, Ap
8	103.5	2.3	522	4	US-09-075-272-3	Sequence 3, Appli
9	103.5	2.3	622	4	US-09-075-272-2	Sequence 2, Appli
10	95.5	2.1	3135	1	US-08-323-170B-2	Sequence 2, Appli
11	95.5	2.1	3135	4	US-08-954-441-2	Sequence 2, Appli
12	95	2.1	749	1	US-08-276-099A-15	Sequence 15, Appl
13	95	2.1	749	1	US-08-781-890-15	Sequence 15, Appl
14	95	2.1	749	3	US-09-012-710-9	Sequence 9, Appli
15	95	2.1	749	4	US-09-364-970-4	Sequence 4, Appli
16	95	2.1	749	4	US-09-556-273-9	Sequence 9, Appli
17	95	2.1	749	4	US-08-956-653A-10	Sequence 10, Appl
18	95	2.1	789	4	US-08-971-188-9	Sequence 9, Appli
19	95	2.1	789	4	US-09-374-454-22	Sequence 22, Appl
20	95	2.1	804	1	US-08-785-428-2	Sequence 2, Appli
21	95	2.1	804	2	US-08-996-797-2	Sequence 2, Appli
22	95	2.1	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
23	94.5	2.1	805	4	US-09-134-001C-4821	Sequence 4821, Ap
24	94.5	2.1	1093	4	US-09-315-793-52	Sequence 52, Appl
25	94.5	2.1	2071	4	US-09-415-522-6	Sequence 6, Appli
26	94.5	2.1	3443	2	US-08-416-603-2	Sequence 2, Appli
27	94	2.1	1151	4	US-09-134-001C-3242	Sequence 3242, Ap
28	94	2.1	6095	4	US-09-144-085-2	Sequence 2, Appli
29	93	2.1	534	4	US-09-198-603C-6	Sequence 6, Appli
30	93	2.1	2763	3	US-08-496-944-2	Sequence 2, Appli
31	92.5	2.0	596	4	US-09-752-165-2	Sequence 2, Appli
32	92.5	2.0	1162	2	US-08-663-566A-15	Sequence 15, Appl
33	92.5	2.0	1162	2	US-08-023-610-15	Sequence 15, Appl
34	92.5	2.0	1162	2	US-08-288-065A-15	Sequence 15, Appl
35	92.5	2.0	1162	2	US-08-362-240A-15	Sequence 15, Appl
36	92.5	2.0	1162	5	PCT-US95-10245-15	Sequence 15, Appl
37	92	2.0	600	4	US-09-134-001C-3636	Sequence 3636, Ap
38	92	2.0	1030	4	US-09-091-117-2	Sequence 2, Appli
39	91.5	2.0	476	4	US-09-134-001C-4458	Sequence 4458, Ap
40	91.5	2.0	3169	4	US-09-453-702B-257	Sequence 257, App
41	90.5	2.0	524	2	US-08-615-942A-2	Sequence 2, Appli
42	90.5	2.0	538	4	US-09-309-572-12	Sequence 12, Appl
43	90.5	2.0	648	4	US-09-134-001C-5161	Sequence 5161, Ap
44	90.5	2.0	1737	4	US-09-309-572-13	Sequence 13, Appl
45	90	2.0	1175	2	US-08-598-305A-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-08-416-603-4
; Sequence 4, Application US/08416603
; Patent No. 5866780
; GENERAL INFORMATION:
; APPLICANT: Law, Marcus
; APPLICANT: Hebara, Ledare
; APPLICANT: Reddick, Bradford B.
; TITLE OF INVENTION: Maize Chlorotic Dwarf Virus Genome and
; TITLE OF INVENTION: Uses Therefor
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/416,603
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lloyd, Jeffrey
; REGISTRATION NUMBER: 35,589
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3457 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-416-603-4

Query Match 2.5%; Score 112.5; DB 2; Length 3457;
Best Local Similarity 17.7%; Pred. No. 0.42;
Matches 164; Conservative 121; Mismatches 275; Indels 365; Gaps 45;

```
Qy          52 QRGTAG----GTAKSQGLQISAGFQELSXAERGGRRRDFHDLGVN--TRQNLDHVKESKT 105
               |  | |     ::: : | : ||:  ||  :  | | ||| : |:   |: :|
Db           8 QNPTQGSIPENSSQDRNLGVPAGY-SLSVEDPFGNRSDFH-IPVHQIIREEIDRPN---- 61

Qy         106 GSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDG 165
                     | : :| | | |           ||    |: |:::  :|: ::|| :
```

```
Db         62 -----WVPICSNDFHLNSE----------DYCEECESERIKNFEIFRSQNLIDQ------ 100

Qy        166 TILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNI--IFRRLLKIM 223
Db        101 ---HLNLCTDSKDCDHFSCFSTSTSCRF-----------CPFCLFIFNLDKFYKQNLYLI 146

Qy        224 NLQQIGRNYYNP-------------NDPID---IPSHRL------VIWPGFTTSILQ--Y 259
Db        147 SRQALARLFHGSAEELLSRAIFFTYNICIDAEVVANNRIGCEYVKLFHPDLRPSITSPPY 206

Qy        260 ENSIMLCTDVSHKV---------------LRSETV------------------------ 279
Db        207 ASDWVMCDNAKHLFECLGLGDTTRGHLYGLISENAYWNATCSKCGACCQGANARTAIPIV 266

Qy        280 -------LD------FMFNFYHQTEEHKFQEQ-------------------VSKEXIGLV 307
Db        267 MALQYCRVDVYYSEYYLYHIYAPEERMKIDQQTAHLLHSIIRGAPAVDCSELSQEPIHRM 326

Qy        308 VL---------------------------------TKYNIKT------------------ 316
Db        327 VMDSSKLVALDSTIRHPKSQGSLLDSECDHEFILRTSHGIKIPMSKSLFISFLTMGAYHG 386

Qy        317 YRVDDIDWDQNPKSTFKKADGSGV----SFLEYYRKQYNQEITDLKQPVLVSQPKRRRGP 372
Db        387 YAHDD-QQEQNAIISFGGMPGVNLACNKNFLRMHKLFYSGSFR--RRPLFMSQ------- 436

Qy        373 GGTLPGPAMLIPELCYLTGLTDKMRNDFN------VMKDLAVHTRL-TPEQRQREVGRLI 425
Db        437 ----------IPS----TNAT--AQSGFNDEEFERLMAEEGVHVKVERPIAERFDYEDVI 480

Qy        426 DYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFA--------D 477
Db        481 DIYDETDHDRTRALGLGQVFGGLLKGISHCV---DSLH---KVFDFPLDLAIEAAQKTGD 534

Qy        478 WSK-------ETR---GAPLI--------------------SVKPLDNWLLIYTRRN--- 504
Db        535 WLEGNKAAVDETKICVGCPEIQKDMISFQNETKEAFELIRSSIKKLSEGIDKITKMNATN 594

Qy        505 -------YEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVV 557
Db        595 FERILDGIKPIESRLTELENKAPASDSKAMEALVQAVKD-----LKIMKE---------- 639

Qy        558 CLLSSNRK-DKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGEL 616
Db        640 AMLDLNRRLSKLEGKK----SDGQT-----TEGTAGEQQPIPKTPTRVKARPVVKQSGTI 690

Qy        617 WRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLK 676
Db        691 --------------------------------MVNEESTE-----TFRDNESRVTDPNR 712

Qy        677 VCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGR-----GY 731
Db        713 SDMFAAVTA-----EYL---VKSFTWKVSDGQDKVLADLDLPQ--DLWKSNSRLSDIMGY 762

Qy        732 NPRXTVIVVKKRVNTRFFAQSGGRL 756
Db        763 FQYYDATGITFRITTTCVPMHGGTL 787


RESULT 2
US-09-224-024-28
; Sequence 28, Application US/09224024
; Patent No. 6056953
;  GENERAL INFORMATION:
;    APPLICANT:  Leslie Hickle
;    APPLICANT:  Jewel Payne
;    TITLE OF INVENTION:  Materials and Methods for the Control of
;    TITLE OF INVENTION:  Calliphoridae Pests
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David R. Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/224,024
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/856,226
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, David R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  MA79
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  904-375-8100
;      TELEFAX:  904-372-5800
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1180 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-224-024-28

  Query Match             2.5%;  Score 112;  DB 3;  Length 1180;
  Best Local Similarity   18.1%;  Pred. No. 0.07;
  Matches  156;  Conservative  114;  Mismatches  274;  Indels  316;  Gaps   39;

Qy        186 RNGEDVRITITLT-----NELP------PTSPTCLQFYNIIF--------RRLLKIMNLQ 226
Db        165 QNTQDVRTQIQLVHYHFQNVIPELVNSCPPNPSDCDYYNILVLSSYAQAANLHLTVLNQA 224

Qy        227 QIGRNYYNPNDPID----IPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLR-----SE 277
Db        225 VKFEAYLKNNRQFDYLEPLPT-AIDYYPVLTKAIEDYTN---YCVTTYKKGLNLIKTTPD 280

Qy        278 TVLDFMFNFYHQTEEHKFQEQVSKEXIGLVVL-TKYNIKTYRV-----------DDIDWD 325
Db        281 SNLDGNINW---NTYNTYRTKMTTAVLDLVALFPNYDVGKYPIGVQSELTREIYQVLNFE 337

Qy        326 QNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPE 385
Db        338 ESP-------------YKYYDFQY-QEDSLTRRPHLFT----------------WLDS 365

Qy        386 LCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLS- 444
Db        366 LNFYEKAQTTPNNFFTSHYNMFHYTLDNISQKSSVFG--------NHNVTDKLKSLGLAT 417

Qy        445 ----FDSNLLSFSG-------------------RILQTEKIHQGGK-TFDYNPQFADWS 479
Db        418 NIYIFLLNVISLDNKYLNDYNNISKMDFFITNGTRLLEKELTAGSGQITYDVN------- 470

Qy        480 KETRGAPLISVKPLDNW---LLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDD 536
Db        471 KNIFGLPIL--KPRENQAIPTLFPTYDNYSHILSFIKSL--SIPA--------------- 511

Qy        537 RTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQT 596
Db        512 -------------TYKTQVYTFAWTHSSVDPKNTIYTHLTTQIP----AVKANSLG---- 550

Qy        597 VMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYH--------------DMTAGR 641
Db        551 ---TASKV-VQGPGHTGGDLIDFKDHFK------ITCQHSNFQQSYFIRIRYASNGSANT 600

Qy        642 RSIAGF----VASI---------------------------------NEGMTRWFSRCI 663
```

```
Db        601 RAVINLSIPGVAELGMALNPTFSGTDYTNLKYKDFQYLEFSNEVKFAPNQNISLVFNRSD 660

Qy        664 FQDRGQELVDGLKVCLQAALRAWNSCNEYMP-----------------SRII--VYRDGV 704

Db        661 VYTNTTVLIDKI--------------EFLPITRSIREDREKQKLETVQQIINTFYANPI 705

Qy        705 GDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQN------ 758

Db        706 KNTLQSELTDYDIDQAANLVECISEELYPKEKMLLLDEVKNAKQLSQSRNVLQNGDFESA 765

Qy        759 PLPGTVID---VEVTRPEWYDFFIVSQAVR--SGSVSPTHYNVIYDNSGLKP-------- 805

Db        766 TLGWTTSDNITIQEDDPIFKGHYLHMSGAREIDGTIFPTYIFQKIDESKLKPYTRYLVRG 825

Qy        806 ------------------DHIQRLTYKLCHIY---YNWPGVIR-----VPA-------- 830

Db        826 FVGSSKDVELVVSRYGEEIDAIMNVPADLNYLYPSTFDCEGSNRCETSAVPANIGNTSDM 885

Qy        831 --PCQYAHKLAFLVGQSIHR 848

Db        886 SYSCQYDTGKKHVVCQDSHQ 905
```

RESULT 3
PCT-US94-07902-28
; Sequence 28, Application PC/TUS9407902
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT: Street address: 4980 Carroll Canyon Road
; APPLICANT: City: San Diego
; APPLICANT: State/Province: California
; APPLICANT: Country: US
; APPLICANT: Postal code/Zip: 92121
; APPLICANT: Phone number: (619) 453-8030 Fax number: . (619)453-6991
; APPLICANT: Telex number:
; TITLE OF INVENTION: Materials and Methods for the Control of
; TITLE OF INVENTION: Calliphoridae Pests
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07902
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA79
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1180 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-07902-28

Query Match 2.5%; Score 112; DB 5; Length 1180;
Best Local Similarity 18.1%; Pred. No. 0.07;
Matches 156; Conservative 114; Mismatches 274; Indels 316; Gaps 39;

```
Qy        186 RNGEDVRITITLT-----NELP------PTSPTCLQFYNIIF--------RRLLKIMNLQ 226

Db        165 QNTQDVRTQIQLVHYHFQNVIPELVNSCPPNPSDCDYYNILVLSSYAQAANLHLTVLNQA 224

Qy        227 QIGRNYYNPNDPID----IPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLR-----SE 277

Db        225 VKFEAYLKNNRQFDYLEPLPT-AIDYYPVLTKAIEDYTN---YCVTTYKKGLNLIKTTPD 280

Qy        278 TVLDFMFNFYHQTEEHKFQEQVSKEXIGLVVL-TKYNIKTYRV-----------DDIDWD 325

Db        281 SNLDGNINW---NTYNTYRTKMTTAVLDLVALFPNYDVGKYPIGVQSELTREIYQVLNFE 337

Qy        326 QNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPE 385

Db        338 ESP--------------YKYYDFQY-QEDSLTRRPHLFT----------------WLDS 365

Qy        386 LCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLS- 444

Db        366 LNFYEKAQTTPNNFFTSHYNMFHYTLDNISQKSSVFG--------NHNVTDKLKSLGLAT 417

Qy        445 ----FDSNLLSFSG-------------------RILQTEKIHQGGK-TFDYNPQFADWS 479

Db        418 NIYIFLLNVISLDNKYLNDYNNISKMDFFITNGTRLLEKELTAGSGQITYDVN------- 470

Qy        480 KETRGAPLISVKPLDNW---LLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDD 536

Db        471 KNIFGLPIL--KPRENQAIPTLFPTYDNYSHILSFIKSL--SIPA--------------- 511

Qy        537 RTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQT 596

Db        512 ------------TYKTQVYTFAWTHSSVDPKNTIYTHLTTQIP----AVKANSLG---- 550

Qy        597 VMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYH--------------DMTAGR 641

Db        551 ---TASKV-VQGPGHTGGDLIDFKDHFK------ITCQHSNFQQSYFIRIRYASNGSANT 600

Qy        642 RSIAGF----VASI---------------------------------NEGMTRWFSRCI 663

Db        601 RAVINLSIPGVAELGMALNPTFSGTDYTNLKYKDFQYLEFSNEVKFAPNQNISLVFNRSD 660

Qy        664 FQDRGQELVDGLKVCLQAALRAWNSCNEYMP-----------------SRII--VYRDGV 704

Db        661 VYTNTTVLIDKI--------------EFLPITRSIREDREKQKLETVQQIINTFYANPI 705

Qy        705 GDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQN------ 758

Db        706 KNTLQSELTDYDIDQAANLVECISEELYPKEKMLLLDEVKNAKQLSQSRNVLQNGDFESA 765

Qy        759 PLPGTVID---VEVTRPEWYDFFIVSQAVR--SGSVSPTHYNVIYDNSGLKP-------- 805

Db        766 TLGWTTSDNITIQEDDPIFKGHYLHMSGAREIDGTIFPTYIFQKIDESKLKPYTRYLVRG 825

Qy        806 ------------------DHIQRLTYKLCHIY---YNWPGVIR-----VPA-------- 830

Db        826 FVGSSKDVELVVSRYGEEIDAIMNVPADLNYLYPSTFDCEGSNRCETSAVPANIGNTSDM 885

Qy        831 --PCQYAHKLAFLVGQSIHR 848

Db        886 SYSCQYDTGKKHVVCQDSHQ 905
```

RESULT 4
US-08-756-317-11
; Sequence 11, Application US/08756317
; Patent No. 5849894
; GENERAL INFORMATION:
; APPLICANT: Clemente, Thomas E.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Mitsky, Timothy A.
; APPLICANT: Stark, David M.
; TITLE OF INVENTION: Improved Rhodospirillum Rubrum
; TITLE OF INVENTION: Poly-B-Hydroxyalkonoate Synthase

```
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  TX
;      COUNTRY:  USA
;      ZIP:  77210-4433
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/756,317
;      FILING DATE:  25-NOV-1996
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/007,693
;      FILING DATE:  29-NOV-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Patterson, Melinda L.
;      REGISTRATION NUMBER:  33,062
;      REFERENCE/DOCKET NUMBER:  MOBT:008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (713) 787-1400
;      TELEFAX:  (713) 787-1440
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  560 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-756-317-11

  Query Match             2.4%;  Score 110.5;  DB 2;  Length 560;
  Best Local Similarity   19.6%;  Pred. No. 0.027;
  Matches  119;  Conservative   79;  Mismatches  232;  Indels  177;  Gaps   27;

Qy        250 PGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQE---------QVS 300
                  |    : :  :|:|:          ||   ::  : |   |:  |           |:
Db         10 PTLPATSMNVQNAIL--------GLRGRDLISTLRNVSRQSLRHPLHTAHHLLALGGQLG 61

Qy        301 KEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQP 360
              :  :|    | : | :  |  |  | |||                 :||:     :   ||
Db         62 RVILGDTPL-QPNPRDPRFSDPTWSQNP----------------FYRRGLQAYLAWQKQT 104

Qy        361 VL------VSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTP 414
               |      :    | |         |  :  |     : | :    :::  |||
Db        105 RLWIEESHLDDDDRAR---------AHFLFNL-----INDALAPSNSLLNPLAVKELFNS 150

Qy        415 --EQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYN 472
                :   | |  |:| :  || : |:: :       || : :| ::   :: :     |
Db        151 GGQSLVRGVAHLLDDLRHNDGLPRQVDERAFEVGGNLAATAGAVVFRNELLE---LIQYK 207

Qy        473 PQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMI 532
              |     |::    ||: | |  |   |:   :  : || :| : |     |:|:
Db        208 PM----SEKQHARPLLVVPPQINKFYIF---DLSSTNSFVQYMLK----NGLQVFMVSWR 256

Qy        533 EVDDR-----TEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVV 587
               | |       :|:: |:: : |   |      | ||              |       |
Db        257 NPDPRHREWGLSSYVQALEEALNACRSI-----SGNR------------DPNLMGACAG 298

Qy        588 ARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGF 647
               |:              ||| : :   :| ||     || :  :|   :  |   |:
Db        299 GLTMA------------ALQGHLQAKHQLRRVRSATYLVSL--LDSKFESPA---SLFAD 341

Qy        648 VASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDG 707
               :|     | : | :        :|| :|   | : ||   |:
Db        342 EQTIEAAKRRSYQRGV--------LDGAEV---ARIFAWMRPND---------------- 374

Qy        708 QLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRL-----QNPLP- 761
                : || |  :|        |  |    |:     :||  |   | |      |||
```

```
Db        375 ---LIWNYWVNNYL-------LGKTPPAFDILYWNADSTRLPAALHGDLLDFFKLNPLTH 424

Qy        762 -------GTVIDVEVTRPEWYDFFIVS----------QAVRSGSVSPTHYNVIYDNSGLK 804
                    || ||::       | | |:            ||  :       :  |||
Db        425 PAGLEVCGTPIDLQKVE---LDSFTVAGSNDHITPWDAVYRSALLLGGDRRFVLANSG-- 479

Qy        805 PDHIQRL 811
                |||  :
Db        480 --HIQSI 484


RESULT 5
US-08-252-966B-12
; Sequence 12, Application US/08252966B
; Patent No. 5624818
;  GENERAL INFORMATION:
;    APPLICANT:  Eisenman, Robert N.
;    APPLICANT:  Hurlin, Peter J.
;    APPLICANT:  Ayer, Donald E.
;    TITLE OF INVENTION:  Regulatory Proteins that Dimerize with
;    TITLE OF INVENTION:  Mad or Max
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson, and KindnessPLLC
;      STREET:  1420 Fifth Ave., Suite 2800
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/252,966B
;      FILING DATE:  01-JUN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Shelton, Dennis K.
;      REGISTRATION NUMBER:  26,997.
;      REFERENCE/DOCKET NUMBER:  FHCR17694
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 682-8100
;      TELEFAX:  (206) 224-0779
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1253 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;      DESCRIPTION:  translation of msina cDNA; see Figure 23
;    HYPOTHETICAL:  YES
;    ORIGINAL SOURCE:
;      ORGANISM:  Mus musculus
US-08-252-966B-12

  Query Match             2.4%;  Score 110;  DB 1;  Length 1253;
  Best Local Similarity   18.7%;  Pred. No. 0.13;
  Matches  121;  Conservative  106;  Mismatches  229;  Indels  192;  Gaps   31;

Qy         70 GFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSR----- 124
               |::|    |   : |    : :     :|:    : |||   :  |    : | |
Db        542 GYKESVHLESFPKERATEGIAM----EIDYASCKRLGSSYRALPKSYQQPKCTGRTPLCK 597

Qy        125 ----------PQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKC--HAFDGTILFLPK 172
                         | |:           |     |:   |:|: | :|    |:  ::
Db        598 EVLNDTWVSFPSWS-----------EDSTFVSSKKTQYEEHIYRCEDERFELDVVLETN 645

Qy        173 RLQQKVTEVFSK--TRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLK--------- 221
                 :| |   |  :|   : :    | | |  ||        :| |: |:
Db        646 LATIRVLEAIQKKLSRLSAEEQAKFRLDNTLGGTS-------EVIHRKALQRIYADKAAD 698
```

```
Qy        222 -IMNLQQIGRNYYNPNDPIDIPSHRLVI----W----PGFTTSILQYENSIMLCTDVSH- 271
Db        699 IIDGLRK------NPSIAVPIVLKRLKMKEEEWREAQRGF-NKVWREQNEKYYLKSLDHQ 751

Qy        272 ---------KVLRSETVLDFMFNFYHQTEEHKFQE----------------QVSKEXIG 305
Db        752 GINFKQNDTKVLRSKSLLNEIESIYDERQEQATEENAGVPVGPHLSLAYEDKQILEDAAA 811

Qy        306 LVVLTKYNIK---------TYRVDDIDWDQNPKSTF-KKADGSGVSFLEYYRKQYNQEIT 355
Db        812 LII---HHVKRQTGIQKEDKYKIKQIMHHFIPDLLFAQRGDLSDV-------EEEEEEEM 861

Qy        356 DLKQPVLVSQPKRRRGPGGTLPGPAMLIPELC--YLTGLTDKMRNDFNVMKDLAVHTRLT 413
Db        862 DVDE--ATGAPKKHNGVGGSPPKSKLLFSNTAAQKLRGM-DEVYNLFYVNNNWYIFMRLH 918

Qy        414 P----------EQRQREVGRLIDYIHKNDNVQRELRDW-----GLSFDSNLLSFSGRILQ 458
Db        919 QILCLRLLRICSQAERQI--------EEENRERE---WEREVLGIKRDK---SDSPAIQL 964

Qy        459 TEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEA--ANSLIQNLF 516
Db        965 RLKEPMDVDVEDYYPAFLDMVRSLLDGNIDSSQYEDSLREMFTIHAYIAFTMDKLIQSIV 1024

Qy        517 K-----VTPAMGMQMRKAIMIE---------VDDRT------EAYLRVLQQKVT------ 550
Db       1025 RQLQHIVSDEVCVQVTDLYLAENNNGATGGQLNSQTSRSLLESAYQRKAEQLMSDENCFK 1084

Qy        551 -------ADTQIVVCLLSSNRKDK---------YDAIKKYLCTDCPTP 582
Db       1085 LMFIQSQGQVQLTVELLDTEEENSDDPVEAERWSDYVERYMSSDTTSP 1132
```

```
RESULT 6
US-08-252-966B-18
; Sequence 18, Application US/08252966B
; Patent No. 5624818
;  GENERAL INFORMATION:
;    APPLICANT:  Eisenman, Robert N.
;    APPLICANT:  Hurlin, Peter J.
;    APPLICANT:  Ayer, Donald E.
;    TITLE OF INVENTION:  Regulatory Proteins that Dimerize with
;    TITLE OF INVENTION:  Mad or Max
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson, and KindnessPLLC
;      STREET:  1420 Fifth Ave., Suite 2800
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/252,966B
;      FILING DATE:  01-JUN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Shelton, Dennis K.
;      REGISTRATION NUMBER:  26,997.
;      REFERENCE/DOCKET NUMBER:  FHCR17694
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 682-8100
;      TELEFAX:  (206) 224-0779
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1261 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;      DESCRIPTION:  translation of msina9 cDNA; see Figure 29A, B, C, D
;    HYPOTHETICAL:  YES
;    ORIGINAL SOURCE:
;      ORGANISM:  Mus musculus
US-08-252-966B-18

  Query Match             2.4%;  Score 110;  DB 1;  Length 1261;
  Best Local Similarity   18.7%;  Pred. No. 0.13;
  Matches  121;  Conservative  106;  Mismatches  229;  Indels  192;  Gaps   31;

Qy         70 GFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSR----- 124
Db        542 GYKESVHLESFPKERATEGIAM----EIDYASCKRLGSSYRALPKSYQQPKCTGRTPLCK 597

Qy        125 ----------PQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKC--HAFDGTILFLPK 172
Db        598 EVLNDTWVSFPSWS-----------EDSTFVSSKKTQYEEHIYRCEDERFELDVVLETN 645

Qy        173 RLQQKVTEVFSK--TRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLK--------- 221
Db        646 LATIRVLEAIQKKLSRLSAEEQAKFRLDNTLGGTS-------EVIHRKALQRIYADKAAD 698

Qy        222 -IMNLQQIGRNYYNPNDPIDIPSHRLVI----W----PGFTTSILQYENSIMLCTDVSH- 271
Db        699 IIDGLRK------NPSIAVPIVLKRLKMKEEEWREAQRGF-NKVWREQNEKYYLKSLDHQ 751

Qy        272 ---------KVLRSETVLDFMFNFYHQTEEHKFQE----------------QVSKEXIG 305
Db        752 GINFKQNDTKVLRSKSLLNEIESIYDERQEQATEENAGVPVGPHLSLAYEDKQILEDAAA 811

Qy        306 LVVLTKYNIK---------TYRVDDIDWDQNPKSTF-KKADGSGVSFLEYYRKQYNQEIT 355
Db        812 LII---HHVKRQTGIQKEDKYKIKQIMHHFIPDLLFAQRGDLSDV-------EEEEEEEM 861

Qy        356 DLKQPVLVSQPKRRRGPGGTLPGPAMLIPELC--YLTGLTDKMRNDFNVMKDLAVHTRLT 413
Db        862 DVDE--ATGAPKKHNGVGGSPPKSKLLFSNTAAQKLRGM-DEVYNLFYVNNNWYIFMRLH 918

Qy        414 P----------EQRQREVGRLIDYIHKNDNVQRELRDW-----GLSFDSNLLSFSGRILQ 458
Db        919 QILCLRLLRICSQAERQI--------EEENRERE---WEREVLGIKRDK---SDSPAIQL 964

Qy        459 TEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEA--ANSLIQNLF 516
Db        965 RLKEPMDVDVEDYYPAFLDMVRSLLDGNIDSSQYEDSLREMFTIHAYIAFTMDKLIQSIV 1024

Qy        517 K-----VTPAMGMQMRKAIMIE---------VDDRT------EAYLRVLQQKVT------ 550
Db       1025 RQLQHIVSDEVCVQVTDLYLAENNNGATGGQLNSQTSRSLLESAYQRKAEQLMSDENCFK 1084

Qy        551 -------ADTQIVVCLLSSNRKDK---------YDAIKKYLCTDCPTP 582
Db       1085 LMFIQSQGQVQLTVELLDTEEENSDDPVEAERWSDYVERYMSSDTTSP 1132
```

```
RESULT 7
US-09-134-001C-3284
; Sequence 3284, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
```

```
; SEQ ID NO 3284
;   LENGTH: 615
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3284

  Query Match             2.4%;  Score 107.5;  DB 4;  Length 615;
  Best Local Similarity   16.8%;  Pred. No. 0.065;
  Matches  130;  Conservative  116;  Mismatches  221;  Indels  307;  Gaps   35;

Qy        114 LSTNHFRLTSR-----PQWALYQYHIDYNPL---------MEARRLRSAL----LFQHED 155
Db         18 ISYNVFRIWRRAIMNHKEWLLADKNIQYRTINAMIKEHIVSEGMRIKEGTCKVEIFLNNH 77

Qy        156 LIG---------KCHAFDGTILFLPKRLQQKV----------TEVFSKTRNGEDVRITIT 196
Db         78 LLSLKVARKSALKRYVFTGDIVLKNKKLSQSIESLEELLQILTEIF-------HINISKR 130

Qy        197 LTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIP-SHRLVIW------ 249
Db        131 LYDELIHSRDSLYETYKHFYNRQTLIHQSMKFSK----------LPDSINFIAWLQHLQD 180

Qy        250 PGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHK-----FQEQVSKEXI 304
Db        181 SGITDD-LSYSESLVIEGHPTHPLTKTKLPL--------STNELKLYAPEFEKVIP---L 228

Qy        305 GLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQPVLVS 364
Db        229 NIMLIEKNHVVTTAIND---DQN--------------FI------LNQVIPEYRDRL--- 262

Qy        365 QPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVM--------------------K 404
Db        263 --------------------KCYLEPLRLNL-NDYRVMLIHPWQYDHTIGEQFEEWIAK 300

Qy        405 DLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLS----------FDSNLLS--F 452
Db        301 KILLPTPFTVESKATLSFRTMDLIHTPYHVKLPVNVQATSAVRTVSSVTTVDGPKLSHAL 360

Qy        453 SGRILQTEKIHQGGKTF--------DYNPQFA----------DWSKETRGAPLISVKPLD 494
Db        361 QGLLQEFPELQVAMEPYGAYAHTASDLSKQLALIIRQKPTIYDYGCTVVTASLVNPNPID 420

Qy        495 NWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTAD-- 552
Db        421 NQAVV-----------------------------------------DSYLKWIENEITLDHI 441

Qy        553 -------TQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIA 605
Db        442 KHFIAIYTQTLVTPLIAYIQNYGIALEAHM-----------------QNTIVNLGPNYK 483

Qy        606 LQMNCK-MGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIF 664
Db        484 MKFIVRDLGGS--RIDLNTLKQKVPQIDVTNE---------SLIADTIEEVIAKFQHAVI 532

Qy        665 QDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLD-- 722
Db        533 QNQLAELI--------------HHFNQY---------DEVIEEEELFEIVREEIEMAIDND 569

Qy        723 -----CLKSIGRGYNPRXTVIV-----------VKKRVNTRFFAQSGGRLQNPL 760
Db        570 KPHAEKLKKILFG----STITVKALLSMRMENKVKKYLNT--------KLDNPI 611


RESULT 8
US-09-075-272-3
; Sequence 3, Application US/09075272
; Patent No. 6136598
;  GENERAL INFORMATION:
;    APPLICANT:  MILLER, A. DUSTY
;    APPLICANT:  WOLGAMOT, GREG
;    APPLICANT:  BONHAM, LYNN
;    TITLE OF INVENTION:  MUS DUNNI ENDOGENOUS RETROVIRAL
;    TITLE OF INVENTION:  PACKAGING CELL LINES
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/075,272
;      FILING DATE:  08-MAY-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/046,140
;      FILING DATE:  09-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poor, Brian W.
;      REGISTRATION NUMBER:  32,928
;      REFERENCE/DOCKET NUMBER:  14538A-003710
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 467-9600
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  522 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-075-272-3

  Query Match             2.3%;  Score 103.5;  DB 4;  Length 522;
  Best Local Similarity   21.2%;  Pred. No. 0.13;
  Matches   76;  Conservative   38;  Mismatches  102;  Indels  143;  Gaps   18;

Qy         27 SQQPGYIQPRPQPPPA--------EGELFGRGRQRGTAGGTAKSQGLQISAGFQELSXAE 78
Db        139 SQPPPY--PLPQQPPAAAPPVAAPQPEPTASGAQ-GPAGGTRSRRG---------RSPAE 186

Qy         79 RGG-------------------------------RRRDFHDLGVNTRQNLDHVKESKTGS 107
Db        187 EGGPDSTVALPLRAHVGGPTPGPNDLIPLQYWPFSSSDLY----NWKTNHPPFSENPSGL 242

Qy        108 SGIIVRLSTNHFRLTSRPQW--------ALYQYHIDYNPLMEARRLRSALLFQHEDLIGK 159
Db        243 TGLLESLMFSH-----QPTWDDCQQLLQVLFTTEERERILMEARK----------NVLGE 287

Qy        160 CHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYN------ 213
Db        288 ----DGT----PTALPNLVDEAFPLNRPNWD--------------------YNTAEGRG 318

Qy        214 --IIFRRLL---------KIMNLQQIGRNYYNPNDPIDIPSHRLV----IWPGFTTSILQ 258
Db        319 RLLVYRRTLVAGLRGAARRPTNLAKVREVLQGQTEPPSVFLERLMEAYRRYTPFDPSSEG 378

Qy        259 YENSIML------CTDVSHKVLRSETVLDFMF--------NFYHQTE-EHKFQEQVSKE 302
Db        379 QKAAVAMAFIGQSAPDIKKKLQRLEGLQDYTLQDLVKEAEKVYHKRETEEERQEREKKE 437


RESULT 9
US-09-075-272-2
; Sequence 2, Application US/09075272
; Patent No. 6136598
;  GENERAL INFORMATION:
;    APPLICANT:  MILLER, A. DUSTY
;    APPLICANT:  WOLGAMOT, GREG
;    APPLICANT:  BONHAM, LYNN
;    TITLE OF INVENTION:  MUS DUNNI ENDOGENOUS RETROVIRAL
```

```
;    TITLE OF INVENTION:  PACKAGING CELL LINES
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/075,272
;      FILING DATE:  08-MAY-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/046,140
;      FILING DATE:  09-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poor, Brian W.
;      REGISTRATION NUMBER:  32,928
;      REFERENCE/DOCKET NUMBER:  14538A-003710
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 467-9600
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  622 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-075-272-2

  Query Match             2.3%;  Score 103.5;  DB 4;  Length 622;
  Best Local Similarity   21.2%;  Pred. No. 0.17;
  Matches   76;  Conservative   38;  Mismatches  102;  Indels  143;  Gaps   18;

Qy         27 SQQPGYIQPRPQPPPA--------EGELFGRGRQRGTAGGTAKSQGLQISAGFQELSXAE 78
              || | |  | || |||        : |    | | | ||||   :|          | ||
Db        239 SQPPPY--PLPQQPPAAAPPVAAPQPEPTASGAQ-GPAGGTRSRRG---------RSPAE 286

Qy         79 RGG-------------------------------RRRDFHDLGVNTRQNLDHVKESKTGS 107
                ||                             | :   | : |     |: :|
Db        287 EGGPDSTVALPLRAHVGGPTPGPNDLIPLQYWPFSSSDLY----NWKTNHPPFSENPSGL 342

Qy        108 SGIIVRLSTNHFRLTSRPQW--------ALYQYHIDYNPLMEARRLRSALLFQHEDLIGK 159
              :|:: |  :|     :| |          |:        |||||:          :::|:
Db        343 TGLLESLMFSH-----QPTWDDCQQLLQVLFTTEERERILMEARK----------NVLGE 387

Qy        160 CHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYN------ 213
                |||    |  |   | | |   |   |                        ||
Db        388 ----DGT----PTALPNLVDEAFPLNRPNWD---------------------YNTAEGRG 418

Qy        214 --IIFRRLL---------KIMNLQQIGRNYYNPNDPIDIPSHRLV----IWPGFTTSILQ 258
                :::|| |         :  || ::         :|  :   ||:     :  |  |
Db        419 RLLVYRRTLVAGLRGAARRPTNLAKVREVLQGQTEPPSVFLERLMEAYRRYTPFDPSSEG 478

Qy        259 YENSIML------CTDVSHKVLRSETVLDFMF--------NFYHQTE-EHKFQEQVSKE 302
                : :: :        |:  |: | | : |:              ||: | | : ||:  ||
Db        479 QKAAVAMAFIGQSAPDIKKKLQRLEGLQDYTLQDLVKEAEKVYHKRETEEERQEREKKE 537


RESULT 10
US-08-323-170B-2
; Sequence 2, Application US/08323170B
; Patent No. 5733772
;  GENERAL INFORMATION:
;    APPLICANT:  Williamson, Kim C.
;    APPLICANT:  Kaslow, David C.
;    TITLE OF INVENTION:  Cloning and Expression of Plasmodium
;    TITLE OF INVENTION:  faliciparum Transmission-Blocking Target Antigen, Pfs230
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/323,170B
;      FILING DATE:  13-OCT-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/010,409
;      FILING DATE:  29-JAN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Quine, Jonathan A.
;      REGISTRATION NUMBER:  P-41,261
;      REFERENCE/DOCKET NUMBER:  015280-113100US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3135 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-323-170B-2

  Query Match             2.1%;  Score 95.5;  DB 1;  Length 3135;
  Best Local Similarity   16.7%;  Pred. No. 21;
  Matches  110;  Conservative  100;  Mismatches  216;  Indels  231;  Gaps   28;

Qy        194 TITLTNELPPTSPTCLQFY-----------NIIFRRLL--------KIMNLQQIGRNYY 233
              | ::| ::||       ||            ||    ||        | | :   :|:
Db       1091 TFSVTFQVPPYIDIKEPFYFMFGCNNNKGEGNIGIVELLISKQEEKIKGCNFHESKLDYF 1150

Qy        234 NPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEH 293
              | |  |   :|          ::  ||| |:                ||   |  :
Db       1151 NEN--ISSDTHE--------CTLHAYENDII-----------------GFNCLETTHPN 1182

Qy        294 KFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQE 353
              : :|    | |      :|                 : :|  :   :: :    : ||
Db       1183 EVEVEVEDAEIYLQPENCFN---------------NVYKGLNSVDITTILKNAQTYN-- 1224

Qy        354 ITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLT 413
              | : | |  :  |                                    :|:::|: :  : |
Db       1225 INNKKTPTFLKIP-------------------------------PYNLLEDVEISCQCT 1252

Qy        414 PEQRQREVGRLIDYIHKNDNV--QRELRDWGLSFDSNLLSFSGRILQTEKIHQG------ 465
              :|  :::  :|    ||| |  :||::      :| |     :| : |  |:
Db       1253 IKQVVKKIKVIIT---KNDTVLLKREVQS-----ESTL---DDKIYKCE--HENFINPRV 1299

Qy        466 GKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFK----VTPA 521
              |||| | ::                                |  |:| |      ||
Db       1300 NKTFDENVEY-------------------------------TCNIKIENFFNYIQIFCPA 1328

Qy        522 MGMQMRKAIMIEVD------------DRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKY 568
              : : | | :  |              :  | :  :   ::   |: :: | :   ::|
Db       1329 KDLGIYKNIQMYYDIVKPTRVPQFKKFNNEELHKLIPNSEMLHKTKEMLILYN---EEKV 1385

Qy        569 DAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMN--CKMGGELWRVDIPLKLV 626
              | :  |:                    : |  |   : |  |     :|:  :  |::
Db       1386 DLLHFYV--------------------FLPIYIKDIYEFNIVCDNSKTMWKNQLGGKVI 1424

Qy        627 MIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFS-------RCIFQDRGQELVDGLKVCL 679
```

```
Db       1425 -------YH-ITVSKREQKVKGCSFDNEHAHMFSYNKTNVKNCIIDAKPKDLIG--FVCP 1474

Qy        680 QAALRAWNSCNEYMP-------SRIIVYRDGVGDGQLKTLVNY-EVPQFLD--------C 723

Db       1475 SGTLKLTNCFKDAIVHTNLTNINGILYLKNNLANFTYKHQFNYMEIPALMDNDISFKCIC 1534

Qy        724 LKSIGRGYNPRX-----TVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWY 775

Db       1535 VDLKKKKYNVKSPLGPKVLRALYKKLNIKFDNYVTGTDQNKYLMTYMDLHLSHKRNY 1591


RESULT 11
US-08-954-441-2
; Sequence 2, Application US/08954441
; Patent No. 6316000
;  GENERAL INFORMATION:
;    APPLICANT:  Williamson, Kim C.
;    APPLICANT:  Kaslow, David C.
;    TITLE OF INVENTION:  Cloning and Expression of Plasmodium
;    TITLE OF INVENTION:  faliciparum Transmission-Blocking Target Antigen, Pfs230
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/954,441
;      FILING DATE:  20-OCT-1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/323,170
;      FILING DATE:  13-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/010,409
;      FILING DATE:  29-JAN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Einhorn, Gregory P.
;      REGISTRATION NUMBER:  38,440
;      REFERENCE/DOCKET NUMBER:  015280-113110US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3135 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-954-441-2

  Query Match             2.1%;  Score 95.5;  DB 4;  Length 3135;
  Best Local Similarity   16.7%;  Pred. No. 21;
  Matches  110;  Conservative  100;  Mismatches  216;  Indels  231;  Gaps   28;

Qy        194 TITLTNELPPTSPTCLQFY-----------NIIFRRLL--------KIMNLQQIGRNYY 233

Db       1091 TFSVTFQVPPYIDIKEPFYFMFGCNNNKGEGNIGIVELLISKQEEKIKGCNFHESKLDYF 1150

Qy        234 NPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEH 293

Db       1151 NEN--ISSDTHE--------CTLHAYENDII-----------------GFNCLETTHPN 1182

Qy        294 KFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQE 353

Db       1183 EVEVEVEDAEIYLQPENCFN---------------NVYKGLNSVDITTILKNAQTYN-- 1224

Qy        354 ITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLT 413

Db       1225 INNKKTPTFLKIP-------------------------------PYNLLEDVEISCQCT 1252

Qy        414 PEQRQREVGRLIDYIHKNDNV--QRELRDWGLSFDSNLLSFSGRILQTEKIHQG------ 465

Db       1253 IKQVVKKIKVIIT---KNDTVLLKREVQS-----ESTL---DDKIYKCE--HENFINPRV 1299

Qy        466 GKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFK----VTPA 521

Db       1300 NKTFDENVEY-------------------------------TCNIKIENFFNYIQIFCPA 1328

Qy        522 MGMQMRKAIMIEVD------------DRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKY 568

Db       1329 KDLGIYKNIQMYYDIVKPTRVPQFKKFNNEELHKLIPNSEMLHKTKEMLILYN---EEKV 1385

Qy        569 DAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMN--CKMGGELWRVDIPLKLV 626

Db       1386 DLLHFYV--------------------FLPIYIKDIYEFNIVCDNSKTMWKNQLGGKVI 1424

Qy        627 MIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFS-------RCIFQDRGQELVDGLKVCL 679

Db       1425 -------YH-ITVSKREQKVKGCSFDNEHAHMFSYNKTNVKNCIIDAKPKDLIG--FVCP 1474

Qy        680 QAALRAWNSCNEYMP-------SRIIVYRDGVGDGQLKTLVNY-EVPQFLD--------C 723

Db       1475 SGTLKLTNCFKDAIVHTNLTNINGILYLKNNLANFTYKHQFNYMEIPALMDNDISFKCIC 1534

Qy        724 LKSIGRGYNPRX-----TVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWY 775

Db       1535 VDLKKKKYNVKSPLGPKVLRALYKKLNIKFDNYVTGTDQNKYLMTYMDLHLSHKRNY 1591


RESULT 12
US-08-276-099A-15
; Sequence 15, Application US/08276099A
; Patent No. 5591825
;  GENERAL INFORMATION:
;    APPLICANT:  McKnight, Steven L
;    APPLICANT:  Hou, Jinzhao
;    TITLE OF INVENTION:  INTERLEUKIN-4 SIGNAL TRANSDUCERS AND
;    TITLE OF INVENTION:  BINDING ASSAYS
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/276,099A
;      FILING DATE:  15-JUL-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard Aron
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59451-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  749 amino acids
;      TYPE:  amino acid
```

```
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-276-099A-15

  Query Match             2.1%;  Score 95;  DB 1;  Length 749;
  Best Local Similarity   18.7%;  Pred. No. 1.9;
  Matches  122;  Conservative   84;  Mismatches  209;  Indels  236;  Gaps   32;

Qy        217 RRLLKIMNLQQI-----GRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVS- 270
              : || | ||::|     |: : |       | |  |:     :: |: |  |:   ::
Db         75 KNLLLIHNLKRIRKVLQGKFHGN-------PMHVAVV----ISNCLREERRILAAANMPI 123

Qy        271 ----HKVLRSETVLDFMFNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQ 326
                  | |:| :| :   |  |:    |   |::::       ||| ::  : |: |:
Db        124 QGPLEKSLQSSSVSERQRNVEHKVSAIKNSVQMTEQD------TKY-LEDLQ-DEFDYRY 175

Qy        327 NPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPEL 386
               | : | : :          |||:  | | :| |   :|:             : ::
Db        176 KTIQTMDQGDKNSI--------LVNQEVLTLLQEMLNSLDFKRK----------EALSKM 217

Qy        387 CYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDW----- 441
              :   || : |                                 :  ||:||
Db        218 TQIVNETDLLMNSM-------------------------------LLEELQDWKKRQQ 244

Qy        442 ----------GLSFDSNLLSFSGRIL-----QTEKIHQGGKTFDYNPQFADWSKETRGAP 486
                       ||    | :     |     | ||: :      |                | |
Db        245 IACIGGPLHNGLDQLQNCFTLLAESLFQLRQQLEKLQEQSTKMTY----------EGDP 293

Qy        487 LISVKPLDNWLLIYTRRNYEAANSLIQNLFKVT------PAMGMQMRKAIMIEVDDRTEA 540
              :    |      |:      | |  || |||| :      | |    :: ::::   :
Db        294 I----PAQRAHLL------ERATFLIYNLFKNSFVVERQPCMPTHPQRPMVLKTLIQFTV 343

Qy        541 YLRVL--------QQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLG 592
               ||:|        | || |     |  ||: |                | |  : : :
Db        344 KLRLLIKLPELNYQVKVKASIDKNVSTLSNRR----------FVLCGTHVKAMSSEESS 392

Qy        593 KQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASIN 652
                  :         :| |  | :              | :  | :|    ||  |   |
Db        393 NGSLSVEFRHLQPKEMKCSTGSK-------------GNEGCHMVTEELHSIT-FETQI- 436

Qy        653 EGMTRWFSRCIFQDRGQELVDGLKVCLQAAL------------RAWNSCNEYMPSRIIV 699
                     |::         || : |: :                || |         ||
Db        437 ---------CLY---------GLTINLETSSLPVVMISNVSQLPNAWAS--------IIW 470

Qy        700 YRDGVGDGQLKTLVNYEVP------QFLDCL-----KSIGRGYN-PRXTVIVVKKRVNTR 747
              |     | |   || : |      | |: :       :||| |  :  ::  |  | :
Db        471 YNVSTNDSQ--NLVFFNNPPSVTLGQLLEVMSWQFSSYVGRGLNSEQLNMLAEKLTVQSN 528

Qy        748 FFAQSGGRL------QNPLPG----------TVIDV--EVTRPEWYDFFIV 780
              :   : | |      :  |||          ::|:  :   | | | :|:
Db        529 Y---NDGHLTWAKFCKEHLPGKTFTFWTWLEAILDLIKKHILPLWIDGYIM 576


RESULT 13
US-08-781-890-15
; Sequence 15, Application US/08781890
; Patent No. 5710266
;  GENERAL INFORMATION:
;    APPLICANT:  McKnight, Steven L
;    APPLICANT:  Hou, Jinzhao
;    TITLE OF INVENTION:  INTERLEUKIN-4 SIGNAL TRANSDUCERS AND
;    TITLE OF INVENTION:  BINDING ASSAYS
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,890
;      FILING DATE:  05-JAN-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/276,099
;      FILING DATE:  15-JUL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard Aron
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59451-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  749 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-781-890-15

  Query Match             2.1%;  Score 95;  DB 1;  Length 749;
  Best Local Similarity   18.7%;  Pred. No. 1.9;
  Matches  122;  Conservative   84;  Mismatches  209;  Indels  236;  Gaps   32;

Qy        217 RRLLKIMNLQQI-----GRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVS- 270
              : || | ||::|     |: : |       | |  |:     :: |: |  |:   ::
Db         75 KNLLLIHNLKRIRKVLQGKFHGN-------PMHVAVV----ISNCLREERRILAAANMPI 123

Qy        271 ----HKVLRSETVLDFMFNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQ 326
                  | |:| :| :   |  |:    |   |::::       ||| ::  : |: |:
Db        124 QGPLEKSLQSSSVSERQRNVEHKVSAIKNSVQMTEQD------TKY-LEDLQ-DEFDYRY 175

Qy        327 NPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPEL 386
               | : | : :          |||:  | | :| |   :|:             : ::
Db        176 KTIQTMDQGDKNSI--------LVNQEVLTLLQEMLNSLDFKRK----------EALSKM 217

Qy        387 CYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDW----- 441
              :   || : |                                 :  ||:||
Db        218 TQIVNETDLLMNSM-------------------------------LLEELQDWKKRQQ 244

Qy        442 ----------GLSFDSNLLSFSGRIL-----QTEKIHQGGKTFDYNPQFADWSKETRGAP 486
                       ||    | :     |     | ||: :      |                | |
Db        245 IACIGGPLHNGLDQLQNCFTLLAESLFQLRQQLEKLQEQSTKMTY----------EGDP 293

Qy        487 LISVKPLDNWLLIYTRRNYEAANSLIQNLFKVT------PAMGMQMRKAIMIEVDDRTEA 540
              :    |      |:      | |  || |||| :      | |    :: ::::   :
Db        294 I----PAQRAHLL------ERATFLIYNLFKNSFVVERQPCMPTHPQRPMVLKTLIQFTV 343

Qy        541 YLRVL--------QQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLG 592
               ||:|        | || |     |  ||: |                | |  : : :
Db        344 KLRLLIKLPELNYQVKVKASIDKNVSTLSNRR----------FVLCGTHVKAMSSEESS 392

Qy        593 KQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASIN 652
                  :         :| |  | :              | :  | :|    ||  |   |
Db        393 NGSLSVEFRHLQPKEMKCSTGSK-------------GNEGCHMVTEELHSIT-FETQI- 436

Qy        653 EGMTRWFSRCIFQDRGQELVDGLKVCLQAAL------------RAWNSCNEYMPSRIIV 699
                     |::         || : |: :                || |         ||
Db        437 ---------CLY---------GLTINLETSSLPVVMISNVSQLPNAWAS--------IIW 470

Qy        700 YRDGVGDGQLKTLVNYEVP------QFLDCL-----KSIGRGYN-PRXTVIVVKKRVNTR 747
              |     | |   || : |      | |: :       :||| |  :  ::  |  | :
Db        471 YNVSTNDSQ--NLVFFNNPPSVTLGQLLEVMSWQFSSYVGRGLNSEQLNMLAEKLTVQSN 528

Qy        748 FFAQSGGRL------QNPLPG----------TVIDV--EVTRPEWYDFFIV 780
              :   : | |      :  |||          ::|:  :   | | | :|:
Db        529 Y---NDGHLTWAKFCKEHLPGKTFTFWTWLEAILDLIKKHILPLWIDGYIM 576
```

```
Db        529 Y---NDGHLTWAKFCKEHLPGKTFTFWTWLEAILDLIKKHILPLWIDGYIM 576


RESULT 14
US-09-012-710-9
; Sequence 9, Application US/09012710
; Patent No. 6087478
;  GENERAL INFORMATION:
;    APPLICANT:  Vinkemeier, Uwe
;    APPLICANT:  Moarefi, Ismail
;    APPLICANT:  Darnell, Jr., James E.
;    APPLICANT:  Kuriyan, John
;    TITLE OF INVENTION:  A CRYSTAL OF THE N-TERMINAL DOMAIN OF A
;    TITLE OF INVENTION:  STAT PROTEIN AND METHODS OF USE THEREOF
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Klauber & Jackson
;      STREET:  411 Hackensack Avenue, 4th Floor
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/012,710
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-194
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-487-5800
;      TELEFAX:  201-343-1684
;      TELEX:  133521
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  749 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
US-09-012-710-9

  Query Match             2.1%;  Score 95;  DB 3;  Length 749;
  Best Local Similarity   18.7%;  Pred. No. 1.9;
  Matches  122;  Conservative   84;  Mismatches  209;  Indels  236;  Gaps   32;

Qy        217 RRLLKIMNLQQI-----GRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVS- 270
Db         75 KNLLLIHNLKRIRKVLQGKFHGN-------PMHVAVV----ISNCLREERRILAAANMPI 123

Qy        271 ----HKVLRSETVLDFMFNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQ 326
Db        124 QGPLEKSLQSSSVSERQRNVEHKVSAIKNSVQMTEQD------TKY-LEDLQ-DEFDYRY 175

Qy        327 NPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPEL 386
Db        176 KTIQTMDQGDKNSI--------LVNQEVLTLLQEMLNSLDFKRK----------EALSKM 217

Qy        387 CYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDW----- 441
Db        218 TQIVNETDLLMNSM--------------------------------LLEELQDWKKRQQ 244

Qy        442 ----------GLSFDSNLLSFSGRIL-----QTEKIHQGGKTFDYNPQFADWSKETRGAP 486
Db        245 IACIGGPLHNGLDQLQNCFTLLAESLFQLRQQLEKLQEQSTKMTY----------EGDP 293

Qy        487 LISVKPLDNWLLIYTRRNYEAANSLIQNLFKVT------PAMGMQMRKAIMIEVDDRTEA 540
Db        294 I----PAQRAHLL------ERATFLIYNLFKNSFVVERQPCMPTHPQRPMVLKTLIQFTV 343

Qy        541 YLRVL--------QQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLG 592
Db        344 KLRLLIKLPELNYQVKVKASIDKNVSTLSNRR----------FVLCGTHVKAMSSEESS 392

Qy        593 KQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASIN 652
Db        393 NGSLSVEFRHLQPKEMKCSTGSK-------------GNEGCHMVTEELHSIT-FETQI- 436

Qy        653 EGMTRWFSRCIFQDRGQELVDGLKVCLQAAL------------RAWNSCNEYMPSRIIV 699
Db        437 ---------CLY---------GLTINLETSSLPVVMISNVSQLPNAWAS--------IIW 470

Qy        700 YRDGVGDGQLKTLVNYEVP------QFLDCL-----KSIGRGYN-PRXTVIVVKKRVNTR 747
Db        471 YNVSTNDSQ--NLVFFNNPPSVTLGQLLEVMSWQFSSYVGRGLNSEQLNMLAEKLTVQSN 528

Qy        748 FFAQSGGRL------QNPLPG----------TVIDV--EVTRPEWYDFFIV 780
Db        529 Y---NDGHLTWAKFCKEHLPGKTFTFWTWLEAILDLIKKHILPLWIDGYIM 576


RESULT 15
US-09-364-970-4
; Sequence 4, Application US/09364970
; Patent No. 6235873
; GENERAL INFORMATION:
;  APPLICANT: Bromberg, Jacqueline
;  TITLE OF INVENTION: CONSTITUTIVELY ACTIVE TRANSCRIPTION FACTORS AND THEIR
;  TITLE OF INVENTION: USES FOR IDENTIFYING MODULATORS OF ACTIVITY INCLUDING
;  TITLE OF INVENTION: DYSPROLIFERATIVE CELLULAR CHANGES
;  FILE REFERENCE: 600-1-252
;  CURRENT APPLICATION NUMBER: US/09/364,970
;  CURRENT FILING DATE: 1999-07-31
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 749
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-364-970-4

  Query Match             2.1%;  Score 95;  DB 4;  Length 749;
  Best Local Similarity   18.7%;  Pred. No. 1.9;
  Matches  122;  Conservative   84;  Mismatches  209;  Indels  236;  Gaps   32;

Qy        217 RRLLKIMNLQQI-----GRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVS- 270
Db         75 KNLLLIHNLKRIRKVLQGKFHGN-------PMHVAVV----ISNCLREERRILAAANMPI 123

Qy        271 ----HKVLRSETVLDFMFNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQ 326
Db        124 QGPLEKSLQSSSVSERQRNVEHKVSAIKNSVQMTEQD------TKY-LEDLQ-DEFDYRY 175

Qy        327 NPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPEL 386
Db        176 KTIQTMDQGDKNSI--------LVNQEVLTLLQEMLNSLDFKRK----------EALSKM 217

Qy        387 CYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDW----- 441
Db        218 TQIVNETDLLMNSM--------------------------------LLEELQDWKKRQQ 244

Qy        442 ----------GLSFDSNLLSFSGRIL-----QTEKIHQGGKTFDYNPQFADWSKETRGAP 486
Db        245 IACIGGPLHNGLDQLQNCFTLLAESLFQLRQQLEKLQEQSTKMTY----------EGDP 293

Qy        487 LISVKPLDNWLLIYTRRNYEAANSLIQNLFKVT------PAMGMQMRKAIMIEVDDRTEA 540
Db        294 I----PAQRAHLL------ERATFLIYNLFKNSFVVERQPCMPTHPQRPMVLKTLIQFTV 343
```

```
Qy        541 YLRVL--------QQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLG 592
              ||:|         | || |      |  ||: |               | |  : : :
Db        344 KLRLLIKLPELNYQVKVKASIDKNVSTLSNRR-----------FVLCGTHVKAMSSEESS 392

Qy        593 KQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASIN 652
                  :        :| |   | :                | :   | :|    ||  |   |
Db        393 NGSLSVEFRHLQPKEMKCSTGSK--------------GNEGCHMVTEELHSIT-FETQI- 436

Qy        653 EGMTRWFSRCIFQDRGQELVDGLKVCLQAAL-------------RAWNSCNEYMPSRIIV 699
                    |::          || : |: :                 || |         ||
Db        437 ---------CLY---------GLTINLETSSLPVVMISNVSQLPNAWAS--------IIW 470

Qy        700 YRDGVGDGQLKTLVNYEVP------QFLDCL-----KSIGRGYN-PRXTVIVVKKRVNTR 747
              |      | |    || :  |       | |: :        :||| |   :  ::  |  | :
Db        471 YNVSTNDSQ--NLVFFNNPPSVTLGQLLEVMSWQFSSYVGRGLNSEQLNMLAEKLTVQSN 528

Qy        748 FFAQSGGRL------QNPLPG----------TVIDV--EVTRPEWYDFFIV 780
              :   : | |      :  |||           ::|:  :   | | | :|:
Db        529 Y---NDGHLTWAKFCKEHLPGKTFTFWTWLEAILDLIKKHILPLWIDGYIM 576
```

Search completed: July 22, 2003, 07:06:38
Job time : 14.4084 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 07:01:08 ; Search time 22.1458 Seconds
(without alignments)
4617.236 Million cell updates/sec

Title: US-09-873-737A-6
Perfect score: 4534
Sequence: 1 MTGRARARARGRARGQETAQ.....VGQSIHREPNLSLSNRLYYL 861

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep1:*
11: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep2:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep3:*
13: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4528	99.9	861	10	US-09-873-737A-6	Sequence 6, Appli
2	4286.5	94.5	862	10	US-09-873-737A-4	Sequence 4, Appli
3	4061	89.6	775	14	US-10-043-774B-2	Sequence 2, Appli
4	2765	61.0	523	11	US-09-998-598-2603	Sequence 2603, Ap
5	2765	61.0	523	15	US-10-029-495-3	Sequence 3, Appli
6	1464	32.3	866	15	US-10-029-495-4	Sequence 4, Appli
7	1423.5	31.4	844	14	US-10-043-774B-13	Sequence 13, Appl
8	1422.5	31.4	844	15	US-10-029-495-7	Sequence 7, Appli
9	1406	31.0	843	10	US-09-873-737A-2	Sequence 2, Appli
10	1365.5	30.1	580	10	US-09-801-574-16	Sequence 16, Appl
11	1319.5	29.1	498	15	US-10-029-495-9	Sequence 9, Appli
12	1291.5	28.5	824	15	US-10-029-495-6	Sequence 6, Appli
13	1144	25.2	722	15	US-10-029-495-5	Sequence 5, Appli
14	652	14.4	130	15	US-10-102-806-685	Sequence 685, App
15	520.5	11.5	1101	15	US-10-174-363-54	Sequence 54, Appl
16	517	11.4	1048	15	US-10-174-363-56	Sequence 56, Appl
17	506.5	11.2	1085	15	US-10-174-363-12	Sequence 12, Appl
18	503.5	11.1	915	15	US-10-174-363-40	Sequence 40, Appl
19	502.5	11.1	1073	15	US-10-174-363-14	Sequence 14, Appl
20	501.5	11.1	1058	15	US-10-174-363-38	Sequence 38, Appl
21	501	11.0	186	15	US-10-029-495-2	Sequence 2, Appli
22	497.5	11.0	969	12	US-09-533-029-106	Sequence 106, App
23	497.5	11.0	969	15	US-10-286-264-102	Sequence 102, App
24	496	10.9	1065	15	US-10-174-363-28	Sequence 28, Appl
25	472.5	10.4	723	15	US-10-174-363-24	Sequence 24, Appl
26	470	10.4	901	15	US-10-174-363-8	Sequence 8, Appli
27	461	10.2	904	15	US-10-174-363-55	Sequence 55, Appl
28	460	10.1	821	15	US-10-174-363-4	Sequence 4, Appli
29	452.5	10.0	757	15	US-10-174-363-10	Sequence 10, Appl
30	440	9.7	1145	11	US-09-866-557A-5	Sequence 5, Appli
31	440	9.7	1145	15	US-10-055-797-5	Sequence 5, Appli
32	434.5	9.6	916	15	US-10-174-363-42	Sequence 42, Appl
33	431.5	9.5	910	15	US-10-174-363-22	Sequence 22, Appl
34	414	9.1	577	15	US-10-174-363-2	Sequence 2, Appli
35	386.5	8.5	919	15	US-10-174-363-16	Sequence 16, Appl
36	383.5	8.5	442	15	US-10-174-363-26	Sequence 26, Appl
37	366	8.1	411	15	US-10-174-363-6	Sequence 6, Appli
38	350	7.7	389	15	US-10-174-363-50	Sequence 50, Appl
39	319.5	7.0	346	15	US-10-174-363-32	Sequence 32, Appl
40	210.5	4.6	71	10	US-09-864-761-48051	Sequence 48051, A
41	145	3.2	55	10	US-09-864-761-38135	Sequence 38135, A
42	142	3.1	193	15	US-10-174-363-44	Sequence 44, Appl
43	136	3.0	67	15	US-10-106-698-5433	Sequence 5433, Ap
44	134.5	3.0	1924	11	US-09-866-557A-2	Sequence 2, Appli
45	134.5	3.0	1924	15	US-10-055-797-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-873-737A-6
; Sequence 6, Application US/09873737A
; Patent No. US20020076797A1
; GENERAL INFORMATION:
; APPLICANT: Duke University
; APPLICANT: Lin, Haifan
; TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
; TITLE OF INVENTION: PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
; FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2
; CURRENT APPLICATION NUMBER: US/09/873,737A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/US99/28764
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/110,901
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 861
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (76)
; OTHER INFORMATION: Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (303)
; OTHER INFORMATION: Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (735)
; OTHER INFORMATION: Xaa=Leu or Ile
US-09-873-737A-6

Query Match 99.9%; Score 4528; DB 10; Length 861;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 861; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRGTAGGTA 60

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRGTAGGTA 60

Qy         61 KSQGLQISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFR 120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 KSQGLQISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFR 120

Qy        121 LTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTE 180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTE 180

Qy        181 VFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPID 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPID 240

Qy        241 IPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVS 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVS 300

Qy        301 KEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQP 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQP 360

Qy        361 VLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQRE 420
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQRE 420

Qy        421 VGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSK 480
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 VGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSK 480

Qy        481 ETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDRTEA 540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDRTEA 540

Qy        541 YLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAI 600
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAI 600

Qy        601 ATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFS 660
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFS 660

Qy        661 RCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQF 720
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQF 720

Qy        721 LDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIV 780
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 LDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIV 780

Qy        781 SQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAF 840
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 SQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAF 840

Qy        841 LVGQSIHREPNLSLSNRLYYL 861
                 |||||||||||||||||||
Db        841 LVGQSIHREPNLSLSNRLYYL 861


RESULT 2
US-09-873-737A-4
; Sequence 4, Application US/09873737A
; Patent No. US20020076797A1
; GENERAL INFORMATION:
;  APPLICANT: Duke University
;  APPLICANT:  Lin, Haifan
;  TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
;  TITLE OF INVENTION:  PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
;  FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2
;  CURRENT APPLICATION NUMBER: US/09/873,737A
;  CURRENT FILING DATE:  2001-06-04
;  PRIOR APPLICATION NUMBER: PCT/US99/28764
;  PRIOR FILING DATE: 1999-12-03
;  PRIOR APPLICATION NUMBER: 60/110,901
;  PRIOR FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 862
;   TYPE: PRT
;   ORGANISM: Mus sp.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (90)
;   OTHER INFORMATION: Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (216)
;   OTHER INFORMATION: Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (383)
;   OTHER INFORMATION: Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (816)
;   OTHER INFORMATION: Xaa=Leu or Ile
US-09-873-737A-4

  Query Match             94.5%;  Score 4286.5;  DB 10;  Length 862;
  Best Local Similarity   93.7%;  Pred. No. 0;
  Matches  808;  Conservative   22;  Mismatches   31;  Indels    1;  Gaps    1;

Qy          1 MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRG-TAGGT 59
                ||||||||||||||||||| | ||: |||||||| ||||  | ||:| |||||||   | |
Db          1 MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEGDLVGRGRQRGMVVGAT 60

Qy         60 AKSQGLQISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHF 119
                :||| |||||||||| |||||||||||| ||||||||||||||||||||||:||||||
Db         61 SKSQELQISAGFQELSLAERGGRRRDFHDXGVNTRQNLDHVKESKTGSSGIIVKLSTNHF 120

Qy        120 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVT 179
                |||||||||||||||||||||||||||||||||||:||||||||||||||| |||
Db        121 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVT 180

Qy        180 EVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPI 239
                ||||:||||| |||||||||||||||||||||| ||||||||||||||||||:|||
Db        181 EVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIXFRRLLKIMNLQQIGRNYYNPSDPI 240

Qy        240 DIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQV 299
                |||:||||||||||||||:|||||||||||||||||||| | ||||||||||
Db        241 DIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQV 300

Qy        300 SKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQ 359
                ||| |||:|||||| |||||||||||||||||||||| ||||||||||||||||||||
Db        301 SKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQ 360

Qy        360 PVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 419
                |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        361 PVLVSQPKRRRGPGGTLPGPAMXIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 420

Qy        420 EVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWS 479
                ||||||||||:|||||||||||||||||||||||||||:|||||||||||||||||||
Db        421 EVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWS 480

Qy        480 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDRTE 539
                |||||||||||||||||||||||||||||||||||||||||||:||:||||||||||||
Db        481 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTE 540

Qy        540 AYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 599
                |||| |||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 600

Qy        600 IATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWF 659
                |||||||||||||||||: ||| |||||||||| ||||||||||||||||||||||||
Db        601 IATKIALQMNCKMGGELWRVDMALKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWF 660

Qy        660 SRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQ 719
```

```
                |||:|||||||||||||||||||||||: |||||||||:|||||||||||||||||||||||
Db        661 SRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQ 720

Qy        720 FLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 779
                |||||||:|||||||  ||||||||||  ||||||||||||||||||||||||||||||||
Db        721 FLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 780

Qy        780 VSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLA 839
                |||||||||||||||||||:|||||||||||||| ||:|||||||||||||||||||||
Db        781 VSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKXCHVYYNWPGVIRVPAPCQYAHKLA 840

Qy        840 FLVGQSIHREPNLSLSNRLYYL 861
                |||||||||||||||||||||
Db        841 FLVGQSIHREPNLSLSNRLYYL 862


RESULT 3
US-10-043-774B-2
; Sequence 2, Application US/10043774B
; Publication No. US20030129599A1
; GENERAL INFORMATION:
;  APPLICANT: University of Illinois at Chicago
;  APPLICANT:  Sharma, Arun
;  APPLICANT:  Hoffman, Ronald
;  TITLE OF INVENTION: HUMAN HEMATOPOIETIC GROWTH REGULATORY GENE AND USES
;  FILE REFERENCE: MBHB: CU08/PPA
;  CURRENT APPLICATION NUMBER: US/10/043,774B
;  CURRENT FILING DATE:  2002-06-06
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 775
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-043-774B-2

  Query Match             89.6%;  Score 4061;  DB 14;  Length 775;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches  768;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         90 GVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSAL 149
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 GVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSAL 63

Qy        150 LFQHEDLIGKCHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCL 209
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 LFQHEDLIGKCHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCL 123

Qy        210 QFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDV 269
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 QFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDV 183

Qy        270 SHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPK 329
               |||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||
Db        184 SHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKELIGLVVLTKYNNKTYRVDDIDWDQNPK 243

Qy        330 STFKKADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYL 389
               ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 STFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYL 303

Qy        390 TGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNL 449
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 TGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNL 363

Qy        450 LSFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAAN 509
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 LSFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAAN 423

Qy        510 SLIQNLFKVTPAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYD 569
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 SLIQNLFKVTPAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYD 483

Qy        570 AIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIV 629
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 AIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIV 543

Qy        630 GIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSC 689
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 GIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSC 603

Qy        690 NEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFF 749
               |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        604 NEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVVKKRVNTRFF 663

Qy        750 AQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQ 809
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 AQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQ 723

Qy        810 RLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 861
               ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 RLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 775


RESULT 4
US-09-998-598-2603
; Sequence 2603, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
;  APPLICANT: Stolk, John A.
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Chenault, Ruth A.
;  APPLICANT:  Meagher, Madelein Joy
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.561
;  CURRENT APPLICATION NUMBER: US/09/998,598
;  CURRENT FILING DATE:  2001-11-16
;  NUMBER OF SEQ ID NOS: 2606
;  SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 2603
;   LENGTH: 523
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-998-598-2603

  Query Match             61.0%;  Score 2765;  DB 11;  Length 523;
  Best Local Similarity   99.8%;  Pred. No. 2e-263;
  Matches  522;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        339 GVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRN 398
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRN 60

Qy        399 DFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQ 458
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQ 120

Qy        459 TEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKV 518
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKV 180

Qy        519 TPAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTD 578
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TPAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTD 240

Qy        579 CPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMT 638
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMT 300

Qy        639 AGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRII 698
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRII 360

Qy        699 VYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQN 758
               |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        361 VYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVVKKRVNTRFFAQSGGRLQN 420
```

```
Qy        759 PLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHI 818
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 PLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHI 480

Qy        819 YYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 861
              |||||||||||||||||||||||||||||||||||||||||||
Db        481 YYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 523


RESULT 5
US-10-029-495-3
; Sequence 3, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
;  APPLICANT: Drmanac, Radoje T.
;  APPLICANT:  Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Dickson, Mark C.
;  APPLICANT:  Jones, Lee William
;  APPLICANT:  Ballinger, Dennis G.
;  APPLICANT:  Xue, Aidong
;  APPLICANT:  Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi,Vinod
;  TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
;  FILE REFERENCE: 28110/35905A
;  CURRENT APPLICATION NUMBER: US/10/029,495
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: 09/378,667
;  PRIOR FILING DATE: 1999-08-20
;  PRIOR APPLICATION NUMBER: 09/687,527
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 523
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: hiwi
US-10-029-495-3

  Query Match           61.0%;  Score 2765;  DB 15;  Length 523;
  Best Local Similarity  99.8%;  Pred. No. 2e-263;
  Matches  522;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        339 GVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRN 398
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRN 60

Qy        399 DFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQ 458
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQ 120

Qy        459 TEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKV 518
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKV 180

Qy        519 TPAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTD 578
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TPAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTD 240

Qy        579 CPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMT 638
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMT 300

Qy        639 AGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRII 698
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRII 360

Qy        699 VYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQN 758
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        361 VYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVVKKRVNTRFFAQSGGRLQN 420

Qy        759 PLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHI 818
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 PLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHI 480

Qy        819 YYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 861
              |||||||||||||||||||||||||||||||||||||||||||
Db        481 YYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 523


RESULT 6
US-10-029-495-4
; Sequence 4, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
;  APPLICANT: Drmanac, Radoje T.
;  APPLICANT:  Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Dickson, Mark C.
;  APPLICANT:  Jones, Lee William
;  APPLICANT:  Ballinger, Dennis G.
;  APPLICANT:  Xue, Aidong
;  APPLICANT:  Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi,Vinod
;  TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
;  FILE REFERENCE: 28110/35905A
;  CURRENT APPLICATION NUMBER: US/10/029,495
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: 09/378,667
;  PRIOR FILING DATE: 1999-08-20
;  PRIOR APPLICATION NUMBER: 09/687,527
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 866
;   TYPE: PRT
;   ORGANISM: Drosophila
;   FEATURE:
;   OTHER INFORMATION: aubergine
US-10-029-495-4

  Query Match           32.3%;  Score 1464;  DB 15;  Length 866;
  Best Local Similarity  36.6%;  Pred. No. 9.3e-135;
  Matches  323;  Conservative  165;  Mismatches  340;  Indels   54;  Gaps   17;

Qy          9 ARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRG--TAGGTAKSQGLQ 66
              |||| ||             ::|  ::      |: |:         |   :||       |
Db         10 ARGRGRG-----------RKPNNVEANRGFAPSLGQKSDPSHSEGNQASGGNGGGGDAQ 57

Qy         67 ISAGFQELSXA----------ERGG--RRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRL 114
              :    :: | :           ||    ||   ||  :    :     || |  |  : :
Db         58 VGPSIEKSSLSAVQMHKSEGDPRGSVRGRRLITDLVYSRPPGM----TSKKGVVGTHITV 113

Qy        115 STNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLP--- 171
               |:|::  || | :||| :|: | :|| ||| : |::|: ::|  : |||| :|
Db        114 QANYFKVLKRPNWTIYQYRVDFTPDVEATRLRRSFLYEHKGILGG-YIFDGTNMFCINQF 172

Qy        172 KRLQQK--VTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIG 229
              | :|    | |: :|:|'||:: | |     :  |     |  |:| || :: ::|: :
Db        173 KAVQDSPYVLELVTKSRAGENIEIKIKAVGSVQSTDAEQFQVLNLILRRAMEGLDLKLVS 232

Qy        230 RNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQ 289
              | ||:|   |:: : |: :|||: ||| |:|| |:||::: |||:|:||: : : :
Db        233 RYYYDPQAKINLENFRMQLWPGYQTSIRQHENDILLCSEICHKVMRTETLYNILSDAIRD 292

Qy        290 TEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQ 349
              :::  :|    :  :|:|:|| || ||||:||:|:   |   ||  ||  :|:::||:|:
Db        293 SDD--YQSTFKRAVMGMVILTDYNNKTYRIDDVDFQSTPLCKFKTNDGE-ISYVDYYKKR 349

Qy        350 YNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVH 409
```

```
           ||  | |||||:::|:|  :   ||      |:||||   ||:|| || ||  :: :: |
Db        350 YNIIIRDLKQPLVMSRPTDKNIRGGN-DQAIMIIPELARATGMTDAMRADFRTLRAMSEH 408

Qy        410 TRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTF 469
              ||| |::|   :       :          |: | :  || |:    |:|  |||  | :
Db        409 TRLNPDRRIERLRMFNKRLKSCKQSVETLKSWNIELDSALVEIPARVLPPEKILFGNQKI 468

Qy        470 DYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKA 529
                 |||: | |   :     :: | :|   ||       :|   :   :| | :
Db        469 FVCDARADWTNEFRTCSMFKNVHINRWYVITPSRNLRETQEFVQMCIRTASSMKMNICNP 528

Qy        530 IMIEV-DDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVA 588
              |  |: |||   | : :      | |||: :: |  ::||  |||  | | | ||| |
Db        529 IYEEIPDDRNGTYSQAIDNAAANDPQIVMVVMRSPNEEKYSCIKKRTCVDRPVPSQVVTL 588

Qy        589 RTLGKQQT----VMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSI 644
              : :  :|     :|:|||| :||| |: |  |:| |||  :| || |  |      :|
Db        589 KVIAPRQQKPTGLMSIATKVVIQMNAKLMGAPWQVVIPLHGLMTVGFDVCHSPKNKNKSY 648

Qy        645 AGFVASINEGMT-RWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDG 703
                |||:::  : |:||      :|||| : : | :  |||::   :  :| ||: :|||
Db        649 GAFVATMDQKESFRYFSTVNEHIKGQELSEQMSVNMACALRSYQEQHRSLPERILFFRDG 708

Qy        704 VGDGQLKTLVNYEVPQFLDCL----KSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNP 759
              ||||||  :|| ||    | |    || |:    | | |:| ||:|:|:|  :| |  ||
Db        709 VGDGQLYQVVNSEVNTLKDRLDEIYKSAGKQEGCRMTFIIVSKRINSRYF--TGHR--NP 764

Qy        760 LPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIY 819
              :||||:|  :| || ||||:|||||| |:|||| |||| || ||  | :| |:||: |:|
Db        765 VPGTVVDDVITLPERYDFFLVSQAVRIGTVSPTSYNVISDNMGLNADKLQMLSYKMTHMY 824

Qy        820 YNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 861
              ||: | ||||| | |||||||||| :||:| |:  | |:||:|
Db        825 YNYSGTIRVPAVCHYAHKLAFLVAESINRAPSAGLQNQLYFL 866
```

RESULT 7
US-10-043-774B-13
; Sequence 13, Application US/10043774B
; Publication No. US20030129599A1
; GENERAL INFORMATION:
; APPLICANT: University of Illinois at Chicago
; APPLICANT: Sharma, Arun
; APPLICANT: Hoffman, Ronald
; TITLE OF INVENTION: HUMAN HEMATOPOIETIC GROWTH REGULATORY GENE AND USES
; FILE REFERENCE: MBHB: CU08/PPA
; CURRENT APPLICATION NUMBER: US/10/043,774B
; CURRENT FILING DATE: 2002-06-06
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 844
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: PIWI protein
US-10-043-774B-13

Query Match 31.4%; Score 1423.5; DB 14; Length 844;
Best Local Similarity 36.0%; Pred. No. 8.8e-131;
Matches 310; Conservative 171; Mismatches 312; Indels 67; Gaps 19;

```
Qy         48 GRGRQR-------GTAGGTAKSQGLQISAG--------------------FQELSXAERG 80
              ||||:|        |: |:     :::  |                     :|    | |
Db          6 GRGRRRPLNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG 65

Qy         81 GRRR----DFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDY 136
                   |    | :|  :|||  :: |  || |: |: | | || ||| ::|:| :  ||:::
Db         66 PPERKPWGDQYDY-LNTRP-VELV--SKKGTDGVPVMLQTNFFRLKTKPEWRIVHYHVEF 121

Qy        137 NPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITIT 196
                | :|  |:|  :|  | :|:|  : |||  ||  :: :|::| :  |::   : :|:|
```

```
Db        122 EPSIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIK 181

Qy        197 LTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSI 256
                  :   |  ||  |:| || :| :||: :||| ::|   |:|   :: :|||: |||
Db        182 FVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSI 241

Qy        257 LQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKT 316
              |:|  |:| |:::|||:|:||: | |    |    |  |::|    : |:||| || :|
Db        242 RQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRT 299

Qy        317 YRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTL 376
              ||::|:|: | |||||    |  :||:|||  :||  | |  ||:|:|:  | :
Db        300 YRINDVDFGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNA 357

Qy        377 PGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQ- 435
                :|||||| :|||  :||::| :|: :: :||: |:||     ||  : |:  |
Db        358 SELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPE 414

Qy        436 --RELRDWGLSFDSNLLSFSGRILQTEKI--HQGGKTFDYNPQFADWSKETRGAPLISVK 491
                : |||| :  | |:    |||:  : |  | |      |   ||| : |   :::
Db        415 SVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTP 471

Qy        492 P--LDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRK-AIMIEVDDRTEAYLRVLQQK 548
                || | :|  :||      :|: :|::    ||:::|      |  ||||  |:| :
Db        472 SDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDC 531

Qy        549 VTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQM 608
              | :| ::::||: ::  ::| :|||    |   |:| |  :|  |  ::|:|||||:|:
Db        532 VRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKKPYSLMSIATKIAIQL 591

Qy        609 NCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASIN-EGMTRWFSRCIFQDR 667
              |||:|   | :::||  :| :| |       :|:    :||:: :  : :||
Db        592 NCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSA 651

Qy        668 GQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSI 727
              | : |   :  ||| :   :  :||||: |||||  | || |  :||   :: ||:
Db        652 FDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKT- 710

Qy        728 GRGY------NPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVS 781
              |       |:   ||| : :|||||       ||| |||:|  :| || |||::||
Db        711 --EYARVQLSPPQLAYIVVTRSMNTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVS 764

Qy        782 QAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFL 841
              | || |:|||| |||:| : || |: :|:||||:||:|||| |  |||| |||| ||| |
Db        765 QQVRQGTVSPTSYNVLYSSMGLSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATL 824

Qy        842 VGQSIHREPNLSLSNRLYYL 861
              || ::|  |  :|  : |||
Db        825 VGTNLHSIPQNALEKKFYYL 844
```

RESULT 8
US-10-029-495-7
; Sequence 7, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
; APPLICANT: Drmanac, Radoje T.
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Dickson, Mark C.
; APPLICANT: Jones, Lee William
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Xue, Aidong
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi,Vinod
; TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
; FILE REFERENCE: 28110/35905A
; CURRENT APPLICATION NUMBER: US/10/029,495
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 09/378,667
; PRIOR FILING DATE: 1999-08-20
; PRIOR APPLICATION NUMBER: 09/687,527

```
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 844
;   TYPE: PRT
;   ORGANISM: Drosophila
;   FEATURE:
;   OTHER INFORMATION: piwi
US-10-029-495-7

  Query Match           31.4%;  Score 1422.5;  DB 15;  Length 844;
  Best Local Similarity  35.9%;  Pred. No. 1.1e-130;
  Matches  309;  Conservative  169;  Mismatches  315;  Indels   67;  Gaps   18;

Qy         48 GRGRQR-------GTAGGTAKSQGLQISAG--------------------FQELSXAERG 80
              ||||:|        |: |:     :::  |                     :|    | |
Db          6 GRGRRRPLNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG 65

Qy         81 GRRR----DFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDY 136
                 |    | :|  :|||        || |: |: | | || ||| ::|:| :  ||::: 
Db         66 PTERKPWGDQYDY-LNTR---PAELVSKKGTDGVPVMLQTNFFRLKTKPEWRIVHYHVEF 121

Qy        137 NPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITIT 196
               | :|  |:|  :|  | :|:|  : |||  ||  :: :|:::| :  |::   : :|:| 
Db        122 EPSIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIK 181

Qy        197 LTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSI 256
                 :    |  ||  |:| || :| :||: :||| ::|   |:|   :: :|||: |||
Db        182 FVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSI 241

Qy        257 LQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKT 316
               |:|  |:| |:::|||:|:||: | |    |    |  |::|    : |:||| || :|
Db        242 RQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRT 299

Qy        317 YRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTL 376
               ||::|:|: | |||||    |  :||:|||  :||  | |  ||:|:|:  | :
Db        300 YRINDVDFGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNA 357

Qy        377 PGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQ- 435
                :|||||| :|||  :||::| :|: :: :||: |:||     ||  : |:  |
Db        358 SELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPE 414

Qy        436 --RELRDWGLSFDSNLLSFSGRILQTEKI--HQGGKTFDYNPQFADWSKETRGAPLISVK 491
                : |||| :  | |:    |||:  : |  | |       |   ||| :  |   :::
Db        415 SVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTP 471

Qy        492 P--LDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRK-AIMIEVDDRTEAYLRVLQQK 548
               || | :|  :||     :|: :|::    ||:::|      |  ||||  |:| :
Db        472 SDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDC 531

Qy        549 VTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQM 608
               | :| ::::||: ::  ::| :|||     |   |:| |  :|  |  ::|:||||||:|:
Db        532 VRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTTKKPYSLMSIATKIAIQL 591

Qy        609 NCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASIN-EGMTRWFSRCIFQDR 667
               |||:|   | :::||  :| :| |         :|:    :||:: :  : :||
Db        592 NCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSA 651

Qy        668 GQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSI 727
                | : |   :  ||| :   :  :||||: |||||  | || |  :||   :: ||:
Db        652 FDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKT- 710

Qy        728 GRGY------NPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVS 781
                  |       |:   ||| : :|||||       ||| |||::|  :| || |||::||
Db        711 --EYARVQLSPPQLAYIVVTRSMNTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVS 764

Qy        782 QAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFL 841
               | || |:|||| |||:| : || |: :|:||||:||:|||| |  |||| |||| ||| |
Db        765 QQVRQGTVSPTSYNVLYSSMGLSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATL 824

Qy        842 VGQSIHREPNLSLSNRLYYL 861
               || ::|  |  :|  : |||
Db        825 VGTNLHSIPQNALEKKFYYL 844


RESULT 9
US-09-873-737A-2
; Sequence 2, Application US/09873737A
; Patent No. US20020076797A1
; GENERAL INFORMATION:
;  APPLICANT: Duke University
;  APPLICANT:  Lin, Haifan
; TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
;  TITLE OF INVENTION:  PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
;  FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2
;  CURRENT APPLICATION NUMBER: US/09/873,737A
;  CURRENT FILING DATE:  2001-06-04
;  PRIOR APPLICATION NUMBER: PCT/US99/28764
;  PRIOR FILING DATE: 1999-12-03
;  PRIOR APPLICATION NUMBER: 60/110,901
;  PRIOR FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 843
;   TYPE: PRT
;   ORGANISM: Drosophila sp.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (13)
;   OTHER INFORMATION: Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (106)
;   OTHER INFORMATION: Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (785)
;   OTHER INFORMATION: Xaa=Leu or Ile
US-09-873-737A-2

  Query Match           31.0%;  Score 1406;  DB 10;  Length 843;
  Best Local Similarity  35.8%;  Pred. No. 4.7e-129;
  Matches  308;  Conservative  172;  Mismatches  312;  Indels   68;  Gaps   20;

Qy         48 GRGRQR-------GTAGGTAKSQGLQISAG--------------------FQELSXAERG 80
              ||||:|        |: |:     :::  |                     :|    | |
Db          6 GRGRRRPXNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG 65

Qy         81 GRRR----DFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDY 136
                 |    | :|  :|||  :: |  || |: |: | | || ||  ::|:| :  ||::: 
Db         66 PPERKPWGDQYDY-LNTRP-VELV--SKKGTDGVPVMLQTNFFRXKTKPEWRIVHYHVEF 121

Qy        137 NPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITIT 196
               | :|  |:|  :|  | :|:|  : |||  ||  :: :|:::| :  |::   : :|:| 
Db        122 VPTIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIK 181

Qy        197 LTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSI 256
                 :    |  ||  |:| || :| :||: :||| ::|   |:|   :: :|||: |||
Db        182 FVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSI 241

Qy        257 LQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKT 316
               |:|  |:| |:::|||:|:||: | |    |    |  |::|    : |:||| || :|
Db        242 RQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRT 299

Qy        317 YRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTL 376
               ||::|:|: | |||||    |  :||:|||  :||  | |  ||:|:|:  | :
Db        300 YRINDVDFGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNA 357

Qy        377 PGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQ- 435
                :|||||| :|||  :||::| :|: :: :||: |:||     ||  : |:  |
Db        358 SELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPE 414

Qy        436 --RELRDWGLSFDSNLLSFSGRILQTEKI--HQGGKTFDYNPQFADWSKETRGAPLISVK 491
                : |||| :  | |:    |||:  : |  | |       |   ||| :  |   :::
Db        415 SVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTP 471
```

```
Qy        492 P--LDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRK-AIMIEVDDRTEAYLRVLQQK 548
Db        472 SDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDC 531

Qy        549 VTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQM 608
Db        532 VRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQL 590

Qy        609 NCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASIN-EGMTRWFSRCIFQDR 667
Db        591 NCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSA 650

Qy        668 GQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSI 727
Db        651 FDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKT- 709

Qy        728 GRGY------NPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVS 781
Db        710 --EYARVQLSPPQLAYIVVTRSMNTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVS 763

Qy        782 QAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFL 841
Db        764 QQVRQGTVSPTSYNVLYSSMGXSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATL 823

Qy        842 VGQSIHREPNLSLSNRLYYL 861
Db        824 VGTNLHSIPQNALEKKFYYL 843
```

RESULT 10
US-09-801-574-16
; Sequence 16, Application US/09801574
; Patent No. US20020081592A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Peijing Jeremy
; APPLICANT: Page, David C.
; TITLE OF INVENTION: Reproduction-Specific Genes
; FILE REFERENCE: 0399.2007-002
; CURRENT APPLICATION NUMBER: US/09/801,574
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: 60/187,518
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: 60/261,557
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 580
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-801-574-16

Query Match 30.1%; Score 1365.5; DB 10; Length 580;
Best Local Similarity 43.7%; Pred. No. 2.4e-125;
Matches 255; Conservative 117; Mismatches 203; Indels 9; Gaps 8;

```
Qy        283 MFNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSF 342
Db          1 MHAIYQQNKEH-FQDECSKLLVGSIVITRYNNRTYRIDDVDWNKTPKDSFVMSDGKEITF 59

Qy        343 LEYYRKQYNQEITDLKQPVLVSQP-KRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFN 401
Db         60 LEYYSKNYGITVKEDDQPLLIHRPSERQNNHGMLLKGEILLLPELSFMTGIPEKMKKDFR 119

Qy        402 VMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEK 461
Db        120 AMKDLTQQINLSPKQHHGALECLLQRISQNETASNELTRWGLSLHKDVHKIEGRLLPMER 179

Qy        462 IHQGGKTFDYNPQFADWSKE-TRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTP 520
Db        180 INLRNTSF-VTSEGLNWVKEVTRDASILTI-PMHFWALFYPKRAMDQARELVNMLEKIAG 237

Qy        521 AMGMQMRKAIMIEV-DDRTEAYLRVLQQ--KVTADTQIVVCLLSSNRKDKYDAIKKYLCT 577
```

```
Db        238 PIGMRTSPPAWVELKDDRIETYIRTIQSLLGVEGKIQMVVCIIMGTRDDLYGAIKKLCCV 297

Qy        578 DCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDM 637
Db        298 QSPVPSQVINVRTIGQPTRLRSVAQKILLQMNCKLGGELWGVDIPLKQLMVIGMDVYHDP 357

Qy        638 TAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRI 697
Db        358 SRGMRSVVGFVASINLTLTKWYSRVVFQMPHQEIVDSLKLCLVGSLKKYYEVNHCLPEKI 417

Qy        698 IVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQ 757
Db        418 VVYRDGVSDGQLKTVANYEIPQLQKCFEAFD-NYHPKMVVFVVQKKISTNLYLAAPDHFV 476

Qy        758 NPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCH 817
Db        477 TPSPGTVVDHTITSCEWVDFYLLAHHVRQGCGIPTHYICVLNTANLSPDHMQRLTFKLCH 536

Qy        818 IYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 861
Db        537 MYWNWPGTIRVPAPCKYAHKLAFLSGQILHHEPAIQLCGNLFFL 580
```

RESULT 11
US-10-029-495-9
; Sequence 9, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
; APPLICANT: Drmanac, Radoje T.
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Dickson, Mark C.
; APPLICANT: Jones, Lee William
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Xue, Aidong
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi,Vinod
; TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
; FILE REFERENCE: 28110/35905A
; CURRENT APPLICATION NUMBER: US/10/029,495
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 09/378,667
; PRIOR FILING DATE: 1999-08-20
; PRIOR APPLICATION NUMBER: 09/687,527
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-029-495-9

Query Match 29.1%; Score 1319.5; DB 15; Length 498;
Best Local Similarity 51.2%; Pred. No. 6.5e-121;
Matches 258; Conservative 86; Mismatches 151; Indels 9; Gaps 4;

```
Qy        361 VLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQRE 420
Db          1 MLVSLLKKKRNDNSE-PQLAHLIPELCFLTGLTDQATSDFQLMKAVAEKTRLSPSGRQQR 59

Qy        421 VGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQ---FAD 477
Db         60 LARLVDNIQRNTNARFELETWGLHFGSQ-ISLTGRIVPSEKILMQ----DHICQPVSAAD 114

Qy        478 WSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDR 537
Db        115 WSKDIRTCKILNAQSLNTWLILCSDRTEYVAESFLNCLRRVAGSMGFNVDYPKIIKVQEN 174

Qy        538 TEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTV 597
```

```
Db        175 PAAFVRAIQQYVDPDVQLVMCILPSNQKTYYDSIKKYLSSDCPVPSQCVLARTLNKQGMM 234

Qy        598 MAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTR 657
              |:||||||:|| ||:|||| |:|||| :|:|||    |  :    : | |||:|  :||
Db        235 MSIATKIAMQMTCKLGGELWAVEIPLKSLMVVGIDVCKDALSKDVMVVGCVASVNPRITR 294

Qy        658 WFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEV 717
              |||||| |    :: | ||| :  ||  |   |  :|:|||||| ||||||||||: |||
Db        295 WFSRCILQRTMTDVADCLKVFMTGALNKWYKYNHDLPARIIVYRAGVGDGQLKTLIEYEV 354

Qy        718 PQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDF 777
              || |  :       : | :||||:|:   ||| :    :|||  |||:| | || |||||
Db        355 PQLLSSVAESSSNTSSRLSVIVVRKKCMPRFFTEMNRTVQNPPLGTVVDSEATRNEWYDF 414

Qy        778 FIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHK 837
              :::||    |:||||:||||||::|||||:||||:||||:||||||:: |||||||||
Db        415 YLISQVACRGTVSPTYYNVIYDDNGLKPDHMQRLTFKLCHLYYNWPGIVSVPAPCQYAHK 474

Qy        838 LAFLVGQSIHREPNLSLSNRLYYL 861
              | ||  ||||:||:| |:| |:||
Db        475 LTFLGAQSIHKEPSLELANHLFYL 498


RESULT 12
US-10-029-495-6
; Sequence 6, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
;  APPLICANT: Drmanac, Radoje T.
;  APPLICANT:  Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Dickson, Mark C.
;  APPLICANT:  Jones, Lee William
;  APPLICANT:  Ballinger, Dennis G.
;  APPLICANT:  Xue, Aidong
;  APPLICANT:  Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi,Vinod
;  TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
;  FILE REFERENCE: 28110/35905A
;  CURRENT APPLICATION NUMBER: US/10/029,495
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: 09/378,667
;  PRIOR FILING DATE: 1999-08-20
;  PRIOR APPLICATION NUMBER: 09/687,527
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 824
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
;   FEATURE:
;   OTHER INFORMATION: ciwi2
US-10-029-495-6

  Query Match           28.5%;  Score 1291.5;  DB 15;  Length 824;
  Best Local Similarity   33.8%;  Pred. No. 8.8e-118;
  Matches  289;  Conservative  174;  Mismatches  317;  Indels   75;  Gaps   19;

Qy         48 GRGRQRGTAGGTAKSQGLQISAGFQELSXAERGGRRRDF-----HDLGVNTRQNLDHVKE 102
              | || ||   |:  |                ||: : :      || :  ||      :
Db          4 GSGRGRGRGSGSNNS----------------GGKDQKYLGTIQPDLFI--RQQ----GQ 40

Qy        103 SKTGSSGIIVRLSTNHFRL-TSRPQWALYQYHIDYNPLMEARRLRSALLFQHE--DLIGK 159
              |||||||   :   |   :  ::  :::||||::: | ::::  |  :|  :   | ||
Db         41 SKTGSSGQPQKCFANFIPIEMTQSDYSIYQYHVEFEPTVDSKANREKMLRDNNVTDEIGH 100

Qy        160 CHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRL 219
                 ||| ||:| :  :|          :   : |    ||      |  :  :| | ||
Db        101 HFVFDGMILYLKEEWEQNQMIEVQHPIDRSLICIRFKQTNRFLVDDPQTINIFNTIIRRS 160

Qy        220 LKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETV 279

              : | |:||||:|  |   :| : : | ||: |:|  || : |||  :   |::| |:
Db        161 FDALQLTQLGRNYFNWGDSRAVPDYNMSILPGYETAIRMYEENFMLCVENRFKMVREES- 219

Qy        280 LDFMFNFYH------QTEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFK 333
              |:  :|       |    : ||:::   |  ::|:|| | :|   :|:  :| | |
Db        220 ---MYILFHKELRSCQNNPQRVQEKMNEMYGGTTIITRYNNKLHRYTRLDYSISPLSEFV 276

Qy        334 KADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLT 393
              | ||  ::  ||:: ||  |||   ||:::|: | ::   |  |  : ::|||| ||||
Db        277 K-DGQSITLKEYFKNQYGIEITVDDQPIIISEGKPKQ--PGEPPQVSYIVPELCFPTGLT 333

Qy        394 DKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFS 453
              |:|| || :||::| |||::|:||  |  :||  : ||: |    : ||:|   :| :
Db        334 DEMRKDFKMMKEIAKHTRMSPQQRLVESRKLIVDLSKNEKVMECFKYWGISLGQDLANVQ 393

Qy        454 GRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEA--ANSL 511
               |:|::| : || ||  |  : |:|::  :  :     : ||::|          :
Db        394 ARVLKSEPL-QGKKT--YEGKQAEWARGVKECGIYRGSNMTNWIVIGPGSGNSGLLSQKF 450

Qy        512 IQNLFKVTPAMGMQMRKAIMIEV-----DDRTEAYLRVLQQKVTADTQIVVCLLSSNRKD 566
              |:   ::   : :|: : : : :     :|  |     ::|    |  ::| :|: : |
Db        451 IEEARRLGKILQVQLGEPMCVPIRGISPNDYLEGVKGAIKQVDGEDIHMLVVMLADDNKT 510

Qy        567 KYDAIKKYLCTDCPTPSQCVVARTL-------GKQQTVMAIATKIALQMNCKMGGELWRV 619
              :||::||:|| :|| |:|||  |||       |: : : :|  || ||| || || ||:|
Db        511 RYDSLKKFLCVECPIPNQCVNLRTLAGKSKDGGENKNLGSIVLKIVLQMICKTGGALWKV 570

Qy        620 DIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCL 679
              :|||  |||| | ||| |   :::   |:: :   |:::|:    :  :| : |   :
Db        571 NIPLKNTMIVGYDLYHDSTLKGKTVGACVSTTSNDFTQFYSQTRPHENPTQLGNNLTHFV 630

Qy        680 QAALRAWNSCNEY-MPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGR--------- 729
              : ||: :   |:  :|||:|:|||| ||||:  : | ||    |   ::
Db        631 RKALKQYYDSNDQTLPSRLILYRDGAGDGQIPYIKNTEVKLVRDACDAVTDKAAELSNKV 690

Qy        730 GYNPRXTVIVVKKRVNTRFFAQSGGRLQ---NPLPGTVIDVEVTRPEWYDFFIVSQAVRS 786
                  :   |:| |||| |   | |   |    || ||||:|  |||||  ||::| | |
Db        691 QEKIKLAFIIVTKRVNMRILKQ-GSSLDNAINPQPGTVVDTTVTRPERMDFYLVPQFVNQ 749

Qy        787 GSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSI 846
              |:|:|  ||:|:|:: | ||  |:| :|||||:|||| | :|||||||||||||||  ||:
Db        750 GTVTPVSYNIIHDDTDLGPDKHQQLAFKLCHLYYNWQGTVRVPAPCQYAHKLAFLTAQSL 809

Qy        847 HREPNLSLSNRLYYL 861
              | : |  | ::|::|
Db        810 HDDANGCLRDKLFFL 824


RESULT 13
US-10-029-495-5
; Sequence 5, Application US/10029495
; Publication No. US20030027255A1
; GENERAL INFORMATION:
;  APPLICANT: Drmanac, Radoje T.
;  APPLICANT:  Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Dickson, Mark C.
;  APPLICANT:  Jones, Lee William
;  APPLICANT:  Ballinger, Dennis G.
;  APPLICANT:  Xue, Aidong
;  APPLICANT:  Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi,Vinod
;  TITLE OF INVENTION: STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS
;  FILE REFERENCE: 28110/35905A
;  CURRENT APPLICATION NUMBER: US/10/029,495
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: 09/378,667
;  PRIOR FILING DATE: 1999-08-20
;  PRIOR APPLICATION NUMBER: 09/687,527
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 5
;   LENGTH: 722
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
;   FEATURE:
;   OTHER INFORMATION: ciwi1
US-10-029-495-5

  Query Match             25.2%;  Score 1144;  DB 15;  Length 722;
  Best Local Similarity   33.3%;  Pred. No. 2.5e-103;
  Matches  254;  Conservative  153;  Mismatches  282;  Indels   74;  Gaps   14;

Qy        127 WALYQYHIDYNPLMEARRLRSALLFQHEDL--IGKCHAFDGTILFLPKRLQQKVTEVFSK 184
              :::||||::: | ::::  |  :| |      |||   ||| ||:| :   |
Db          6 YSIYQYHVEFEPTVDSKATRENMLRQPSVTVEIGKHFVFDGMILYLKEEWDQNQMIEVQH 65

Qy        185 TRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSH 244
                :  : |    ||      |  :  :| | ||    | | ||||||::
Db         66 PNDNSLICIRFKKTNRFLVDDPQTINIFNTIIRRSFDAMKLTQIGRNYFD---------- 115

Qy        245 RLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKEXI 304
                      ::||  |      | |||           |    : ||::::
Db        116 -------------WDNSRAL-----RKELRS----------CQNNRQRVQEKMNEVYG 145

Qy        305 GLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQPVLVS 364
              |  ::|:|| | :|   :|  :  | | |:| ||  :   ||:: ||: :||| :| :::|
Db        146 GSTIITRYNNKLHRFTRLDNEITPLSKFQK-DGEQIILKEYFKNQYDIDITDDEQFIIIS 204

Qy        365 QPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRL 424
              : | ::   |  |    ::|||: |||||:|| || :||::| |||::|:|| |  :|
Db        205 EGKPKQ--PGEPPQVNYIVPELCFPTGLTDEMRKDFKMMKEIAKHTRMSPQQRLDETRKL 262

Qy        425 IDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSKETRG 484
              |  : :|  :    : ||:|   :| :   |:|::| : || :   |  : |:|::  :
Db        263 ITKLSQNQTMMECFQYWGISLGQDLANVQARVLKSEPL-QGKR--QYEGKQAEWARGVKE 319

Qy        485 APLISVKPLDNWLLIYTRRNYEA--ANSLIQNLFKVTPAMGMQMRKAIMIEV-----DDR 537
                :     : ||::|       |   |     :   : :|: : : :::     :|
Db        320 CGIYRGSNMTNWIVIGPGSGNSGLLAQKFIAEARNLGRTLQVQLGEPMCVKINGISPNDY 379

Qy        538 TEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTL------ 591
               |     ::     :  ::| :|: : | :||::|||| :|| |:|||  |||
Db        380 LEGLKAAIKSVDGEEIHMLVVMLADDNKTRYDSLKKYLCVECPIPNQCVNLRTLAGKSKD 439

Qy        592 -GKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVAS 650
              |: : : :|  || ||| || || ||:|:|||| |||| | ||| |   :::   |::
Db        440 GGENKNLGSIVLKIVLQMICKTGGALWKVNIPLKSTMIVGYDLYHDSTLKGKTVGACVST 499

Qy        651 INEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNE-YMPSRIIVYRDGVGDGQL 709
              :  |:::|:    :  :| : |   :: :|: :   |:  :|||:|:|||| ||||:
Db        500 TSNDFTQFYSQTRPHENPTQLGNNLTHFVRKSLKQYYDNNDKTLPSRLILYRDGAGDGQI 559

Qy        710 KTLVNYEVPQFLDCLKSIGR---------GYNPRXTVIVVKKRVNTRFFAQ--SGGRLQN 758
                : | ||    |   ::                 :  |:| |||| |   |  |     |
Db        560 PYIKNTEVKLVRDACDAVTDKAAELSNKVQEKIKLAFIIVTKRVNMRILKQGSSSKSAIN 619

Qy        759 PLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHI 818
              | ||||:|  |||||  ||::| | |  |:|:|  ||:|:|::|| ||  |:| :||||:
Db        620 PQPGTVVDTTVTRPERMDFYLVPQFVNQGTVTPVSYNIIHDDTGLGPDKHQQLAFKLCHL 679

Qy        819 YYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 861
              |||| | :|||||||||||||||  ||:| : |  | ::|::|
Db        680 YYNWQGTVRVPAPCQYAHKLAFLTAQSLHDDANGYLRDKLFFL 722


RESULT 14
US-10-102-806-685
; Sequence 685, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA103P1C1
;  CURRENT APPLICATION NUMBER: US/10/102,806
;  CURRENT FILING DATE:  2002-03-22
;  PRIOR APPLICATION NUMBER: 09/925,298
;  PRIOR FILING DATE: 2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05881
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 846
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 685
;   LENGTH: 130
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (61)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-102-806-685

  Query Match             14.4%;  Score 652;  DB 15;  Length 130;
  Best Local Similarity   99.2%;  Pred. No. 4.4e-56;
  Matches  123;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        578 DCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDM 637
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db          5 DCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCXHDM 64

Qy        638 TAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRI 697
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 TAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRI 124

Qy        698 IVYR 701
              ||||
Db        125 IVYR 128


RESULT 15
US-10-174-363-54
; Sequence 54, Application US/10174363
; Publication No. US20030077623A1
; GENERAL INFORMATION:
;  APPLICANT: Butler, Karlene H.
;  APPLICANT:  Cahoon, Rebecca E
;  APPLICANT:  Harvell, Leslie T.
;  APPLICANT:  Rafalski, Antoni J.
;  APPLICANT:  Sakai, Hajime
;  TITLE OF INVENTION: Polynucleotides And Polypeptides Involved In Post-Transcriptio
;  TITLE OF INVENTION:  Gene Silencing
;  FILE REFERENCE: BB1454 US NA
;  CURRENT APPLICATION NUMBER: US/10/174,363
;  CURRENT FILING DATE:  2002-06-17
;  PRIOR APPLICATION NUMBER: 60/298,973
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 56
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 54
;   LENGTH: 1101
;   TYPE: PRT
;   ORGANISM: Oryza sativa
US-10-174-363-54

  Query Match             11.5%;  Score 520.5;  DB 15;  Length 1101;
  Best Local Similarity   22.5%;  Pred. No. 1.6e-41;
  Matches  222;  Conservative  161;  Mismatches  424;  Indels  179;  Gaps   36;

Qy         23 GSTASQQPGYIQPRPQPPPAEGELFGR----------GRQRGTAG--------------- 57
              | : |  ||   |  ||   :|    |          |  ||: |
Db        103 GGSGSHHPGGGPPEYQPRDYQGRGGPRPRGGMPQPYYGGPRGSGGRSVPSGSSRTVPELH 162

Qy         58 ------------------GTAKSQGLQISAG-----FQELSXAERGGRRRDFHDLGVNT 93
                               |::    ::|:|     ||:|:  ::    :       ::
Db        163 QAPHVQYQAPMVSPTPSGAGSSSQPAAEVSSGQVQQQFQQLATRDQSSTSQAIQIAPPSS 222
```

```
Qy         94 RQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQ- 152
Db        223 KSVRFPLRPGK-GTYGDRCIVKANHF-FAELPDKDLHQYDVSITPEVTSRGVNRAVMFEL 280

Qy        153 -----HEDLIGKCHAFDG-----TILFLP----------------------KRLQQKVTE 180
Db        281 VTLYRYSHLGGRLPAYDGRKSLYTAGPLPFASRTFEITLQDEEDSLGGGQGTQRRERLFR 340

Qy        181 VFSKTRNGEDV-RITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPI 239
Db        341 VVIKFAARADLHHLAMFLAGRQADAPQEALQVLDIVLRE-LPTTRYSPVGRSFYSPN--- 396

Qy        240 DIPSHR------LVIWPGFTTSILQYENSIMLCTDVSHKV-LRSETVLDFMFNFYHQ--- 289
Db        397 --LGRRQQLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIDFVAQLLNRDIS 454

Qy        290 ----TEEHKFQEQVSKEXIGLVVLTKYNI-KTYRVDDIDWDQNPKSTFKKAD-GSGVSFL 343
Db        455 VRPLSDSDRVKIKKALRGVKVEVTHRGNMRRKYRISGLTSQATRELSFPVDDRGTVKTVV 514

Qy        344 EYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLT--GLTDKMRNDFN 401
Db        515 QYFLETYGFSIQHTTLPCLQVGNQQR---------PNYLPMEVCKIVEGQRYSKRLNEKQ 565

Qy        402 VMKDLAVHTRLTPEQRQREVGRLIDY--IHKNDNVQRELRDWGLSFDSNLLSFSGRILQT 459
Db        566 ITALLKV-TCQRPQERELDILRTVSHNAYHEDQYAQ----EFGIKIDERLASVEARVLPP 620

Qy        460 E--KIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYE--AANSLIQNL 515
Db        621 PRLKYHDSGREKDVLPRVGQWNMMNK--KMVNGGRVNNWACINFSRNVQDSAARGFCHEL 678

Qy        516 FKVTPAMGMQMR-KAIMIEVDDRTEAYLRVLQQKVT----------ADTQIVVCLLSSNR 564
Db        679 AIMCQISGMDFALEPVLPPLTARPEHVERALKARYQDAMNMLRPQGRELDLLIVILPDNN 738

Qy        565 KDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGE-------LW 617
Db        739 GSLYGDLKRICETDLGLVSQCCLTKHVFKMS--KQYLANVALKINVKVGGRNTVLVDALT 796

Qy        618 RVDIPL---KLVMIVGIDCYHDMTA--GRRSIAGFVASIN-EGMTRWFSRCIFQDRGQEL 671
Db        797 R-RIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDWPEVTKYAGLVSAQAHRQEL 855

Qy        672 VDGL-KV------------CLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVP 718
Db        856 IQDLFKVWQDPHRGTVTGGMIKELLISFKRATGQKPQRIIFYRDGVSEGQFYQVLLYELD 915

Qy        719 QFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGG------RLQNPLPGTVIDVEVTRP 772
Db        916 AIRKACASLEPNYQPPVTFVVVQKRHHTRLFANNHNDQRTVDRSGNILPGTVVDSKICHP 975

Qy        773 EWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPC 832
Db        976 TEFDFYLCSHAGIQGTSRPAHYHVLWDENKFTADELQTLTNNLCYTYARCTRSVSIVPPA 1035

Qy        833 QYAHKLAFLVGQSIHREPNLSLSNRL 858
Db       1036 YYAHLAAFRA--RFYMEPETSDSGSM 1059
```

Search completed: July 22, 2003, 07:07:52
Job time : 23.1458 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 06:25:13 ; Search time 17.1126 Seconds
(without alignments)
4836.881 Million cell updates/sec

Title: US-09-873-737A-6
Perfect score: 4534
Sequence: 1 MTGRARARARGRARGQETAQ.....VGQSIHREPNLSLSNRLYYL 861

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1291.5	28.5	824	2	T20351	hypothetical prote
2	1144	25.2	722	2	T30995	hypothetical prote
3	539.5	11.9	834	2	T41568	argonaute-like pro
4	533	11.8	1014	2	H86438	protein T19E23.7 [
5	510	11.2	988	2	T52134	Zwille protein [im
6	504.5	11.1	997	2	A84678	Argonaute (AGO1)-1
7	500.5	11.0	891	2	T32079	hypothetical prote
8	492	10.9	1000	2	T22391	hypothetical prote
9	476.5	10.5	930	2	A84668	Argonaute (AGO1)-1
10	474	10.5	813	2	JC6569	translation initia
11	430.5	9.5	1032	2	T23164	hypothetical prote
12	430.5	9.5	1035	2	T23165	hypothetical prote
13	422	9.3	887	2	T01113	translation initia
14	421	9.3	958	2	S41013	hypothetical prote
15	421	9.3	1040	2	D88568	protein ZK757.3 [i
16	388	8.6	898	2	T25164	hypothetical prote
17	337.5	7.4	1139	2	T33275	hypothetical prote
18	314.5	6.9	905	2	T23510	hypothetical prote
19	279	6.2	892	2	T15195	hypothetical prote
20	273.5	6.0	945	2	T23965	hypothetical prote
21	252.5	5.6	1057	2	T16676	hypothetical prote
22	242	5.3	965	2	T22933	hypothetical prote
23	235.5	5.2	1032	2	T19324	hypothetical prote
24	234.5	5.2	1030	2	T16114	hypothetical prote
25	227.5	5.0	697	2	T15179	hypothetical prote
26	220	4.9	990	2	F88925	protein T22H9.3 [i
27	218	4.8	324	2	G87774	protein C24A11.3 [
28	214	4.7	363	2	T34486	hypothetical prote
29	213	4.7	939	2	T18974	hypothetical prote
30	210	4.6	881	2	T31818	hypothetical prote
31	208	4.6	674	2	T19268	hypothetical prote
32	195	4.3	567	2	T34339	hypothetical prote
33	171.5	3.8	706	2	H70425	conserved hypothet
34	165.5	3.7	944	2	T18911	hypothetical prote
35	140	3.1	713	2	H64464	hypothetical prote
36	133	2.9	709	2	G96610	probable disease r
37	129	2.8	185	2	T27784	hypothetical prote
38	125	2.8	2035	2	AC0233	yersiniabactin bio
39	125	2.8	2041	2	T17439	peptide synthetase
40	120	2.6	1261	2	S75130	sensory transducti
41	117.5	2.6	659	2	F70175	rep helicase, sing
42	117	2.6	1120	2	H71664	transcription-repa
43	117	2.6	1474	2	T20488	hypothetical prote
44	116	2.6	1839	1	OYBYK	adenylate cyclase
45	115.5	2.5	2035	2	A48654	probable polyketid

ALIGNMENTS

RESULT 1
T20351
hypothetical protein D2030.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T20351
R;Wilkinson, J.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19261
A;Accession: T20351
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-824 <WIL>
A;Cross-references: EMBL:Z73906; PIDN:CAA98113.1; GSPDB:GN00019; CESP:D2030.6
A;Experimental source: clone D2030
C;Genetics:
A;Gene: CESP:D2030.6
A;Map position: 1
A;Introns: 118/3; 223/3; 304/3; 452/1; 573/3; 632/1; 711/2; 773/3

Query Match 28.5%; Score 1291.5; DB 2; Length 824;
Best Local Similarity 33.8%; Pred. No. 4.1e-88;
Matches 289; Conservative 174; Mismatches 317; Indels 75; Gaps 19;

```
Qy         48 GRGRQRGTAGGTAKSQGLQISAGFQELSXAERGGRRRDF-----HDLGVNTRQNLDHVKE 102
                |  ||  ||    |:  |                  ||: : :      || :  ||      :
Db          4 GSGRGRGRGSGSNNS----------------GGKDQKYLGTIQPDLFI--RQQ----GQ 40

Qy        103 SKTGSSGIIVRLSTNHFRL-TSRPQWALYQYHIDYNPLMEARRLRSALLFQHE--DLIGK 159
                |||||||    :   |   :  ::  :::||||::: | :::: |  :|  :   | ||
Db         41 SKTGSSGQPQKCFANFIPIEMTQSDYSIYQYHVEFEPTVDSKANREKMLRDNNVTDEIGH 100

Qy        160 CHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRL 219
                |||  ||:|  :  :|          :   : |    ||      |  :  :| | ||
Db        101 HFVFDGMILYLKEEWEQNQMIEVQHPIDRSLICIRFKQTNRFLVDDPQTINIFNTIIRRS 160

Qy        220 LKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETV 279
                :  | |:||||:|  |   :| : : | ||: |:|  || : ||| :   |::| |:
Db        161 FDALQLTQLGRNYFNWGDSRAVPDYNMSILPGYETAIRMYEENFMLCVENRFKMVREES- 219

Qy        280 LDFMFNFYH------QTEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFK 333
                |:  :|      |    : ||::::   |  ::|:|| | :|   :|:  :| | |
Db        220 ---MYILFHKELRSCQNNPQRVQEKMNEMYGGTTIITRYNNKLHRYTRLDYSISPLSEFV 276

Qy        334 KADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLT 393
                | ||  ::  ||:: ||  |||   ||:::|: | ::   |  |  : ::||||: ||||
Db        277 K-DGQSITLKEYFKNQYGIEITVDDQPIIISEGKPKQ--PGEPPQVSYIVPELCFPTGLT 333

Qy        394 DKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFS 453
                |:|| || :||::| |||::|:||  |  :||  : ||: |    : ||:|   :| :
Db        334 DEMRKDFKMMKEIAKHTRMSPQQRLVESRKLIVDLSKNEKVMECFKYWGISLGQDLANVQ 393
```

```
Qy        454 GRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEA--ANSL 511
Db        394 ARVLKSEPL-QGKKT--YEGKQAEWARGVKECGIYRGSNMTNWIVIGPGSGNSGLLSQKF 450

Qy        512 IQNLFKVTPAMGMQMRKAIMIEV-----DDRTEAYLRVLQQKVTADTQIVVCLLSSNRKD 566
Db        451 IEEARRLGKILQVQLGEPMCVPIRGISPNDYLEGVKGAIKQVDGEDIHMLVVMLADDNKT 510

Qy        567 KYDAIKKYLCTDCPTPSQCVVARTL-------GKQQTVMAIATKIALQMNCKMGGELWRV 619
Db        511 RYDSLKKFLCVECPIPNQCVNLRTLAGKSKDGGENKNLGSIVLKIVLQMICKTGGALWKV 570

Qy        620 DIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCL 679
Db        571 NIPLKNTMIVGYDLYHDSTLKGKTVGACVSTTSNDFTQFYSQTRPHENPTQLGNNLTHFV 630

Qy        680 QAALRAWNSCNEY-MPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGR--------- 729
Db        631 RKALKQYYDSNDQTLPSRLILYRDGAGDGQIPYIKNTEVKLVRDACDAVTDKAAELSNKV 690

Qy        730 GYNPRXTVIVVKKRVNTRFFAQSGGRLQ---NPLPGTVIDVEVTRPEWYDFFIVSQAVRS 786
Db        691 QEKIKLAFIIVTKRVNMRILKQ-GSSLDNAINPQPGTVVDTTVTRPERMDFYLVPQFVNQ 749

Qy        787 GSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSI 846
Db        750 GTVTPVSYNIIHDDTDLGPDKHQQLAFKLCHLYYNWQGTVRVPAPCQYAHKLAFLTAQSL 809

Qy        847 HREPNLSLSNRLYYL 861
Db        810 HDDANGCLRDKLFFL 824


RESULT 2
T30995
hypothetical protein C01G5.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 29-Oct-1999
C;Accession: T30995
R;Bradshaw, H.; Stellyes, L.
submitted to the EMBL Data Library, August 1999
A;Description: The sequence of C. elegans cosmid C01G5.
A;Reference number: Z20956
A;Accession: T30995
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-722 <BRA>
A;Cross-references: EMBL:U50068; PIDN:AAB37734.1
A;Experimental source: strain Bristol N2
C;Genetics:
A;Map position: IV
A;Introns: 58/3; 121/3; 202/3; 350/1; 471/3; 530/1; 609/2; 671/3
A;Note: C01G5.2

  Query Match             25.2%;  Score 1144;  DB 2;  Length 722;
  Best Local Similarity   33.3%;  Pred. No. 3.5e-77;
  Matches  254;  Conservative  153;  Mismatches  282;  Indels   74;  Gaps   14;

Qy        127 WALYQYHIDYNPLMEARRLRSALLFQHEDL--IGKCHAFDGTILFLPKRLQQKVTEVFSK 184
Db          6 YSIYQYHVEFEPTVDSKATRENMLRQPSVTVEIGKHFVFDGMILYLKEEWDQNQMIEVQH 65

Qy        185 TRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSH 244
Db         66 PNDNSLICIRFKKTNRFLVDDPQTINIFNTIIRRSFDAMKLTQIGRNYFD---------- 115

Qy        245 RLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKEXI 304
Db        116 --------------WDNSRAL-----RKELRS----------CQNNRQRVQEKMNEVYG 145

Qy        305 GLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQPVLVS 364
Db        146 GSTIITRYNNKLHRFTRLDNEITPLSKFQK-DGEQIILKEYFKNQYDIDITDDEQFIIIS 204
```

```
Qy        365 QPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRL 424
Db        205 EGKPKQ--PGEPPQVNYIVPELCFPTGLTDEMRKDFKMMKEIAKHTRMSPQQRLDETRKL 262

Qy        425 IDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSKETRG 484
Db        263 ITKLSQNQTMMECFQYWGISLGQDLANVQARVLKSEPL-QGKR--QYEGKQAEWARGVKE 319

Qy        485 APLISVKPLDNWLLIYTRRNYEA--ANSLIQNLFKVTPAMGMQMRKAIMIEV-----DDR 537
Db        320 CGIYRGSNMTNWIVIGPGSGNSGLLAQKFIAEARNLGRTLQVQLGEPMCVKINGISPNDY 379

Qy        538 TEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTL------ 591
Db        380 LEGLKAAIKSVDGEEIHMLVVMLADDNKTRYDSLKKYLCVECPIPNQCVNLRTLAGKSKD 439

Qy        592 -GKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVAS 650
Db        440 GGENKNLGSIVLKIVLQMICKTGGALWKVNIPLKSTMIVGYDLYHDSTLKGKTVGACVST 499

Qy        651 INEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNE-YMPSRIIVYRDGVGDGQL 709
Db        500 TSNDFTQFYSQTRPHENPTQLGNNLTHFVRKSLKQYYDNNDKTLPSRLILYRDGAGDGQI 559

Qy        710 KTLVNYEVPQFLDCLKSIGR---------GYNPRXTVIVVKKRVNTRFFAQ--SGGRLQN 758
Db        560 PYIKNTEVKLVRDACDAVTDKAAELSNKVQEKIKLAFIIVTKRVNMRILKQGSSSKSAIN 619

Qy        759 PLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHI 818
Db        620 PQPGTVVDTTVTRPERMDFYLVPQFVNQGTVTPVSYNIIHDDTGLGPDKHQQLAFKLCHL 679

Qy        819 YYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 861
Db        680 YYNWQGTVRVPAPCQYAHKLAFLTAQSLHDDANGYLRDKLFFL 722


RESULT 3
T41568
argonaute-like protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C;Accession: T41568
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.
submitted to the EMBL Data Library, May 1998
A;Reference number: Z21991
A;Accession: T41568
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-834 <WOO>
A;Cross-references: EMBL:AL023705; PIDN:CAA19275.1; GSPDB:GN00068; SPDB:SPCC736.11
A;Experimental source: strain 972h-; cosmid c736
C;Genetics:
A;Gene: SPDB:SPCC736.11
A;Map position: 3
A;Introns: 43/3
C;Superfamily: rabbit translation initiation factor eIF-2C

  Query Match             11.9%;  Score 539.5;  DB 2;  Length 834;
  Best Local Similarity   22.5%;  Pred. No. 6.8e-32;
  Matches  195;  Conservative  154;  Mismatches  350;  Indels  167;  Gaps   31;

Qy        106 GSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDL--------- 156
Db         16 GGLGKQITLKANFFQIISLPNETINQYHVIVGDGSRVPRKQSQLIWNSKEVKQYFGSSWM 75

Qy        157 ----------IGKCHAFDGTILF------LPKRLQ---QKVTEV-------FSKTRNGED 190
Db         76 NSVYDGRSMCWSKGDIADGTIKVNIGSESHPREIEFSIQKSSKINLHTLSQFVNSKYSSD 135

Qy        191 VRITITLTNELPPTSPTCLQFYNIIFRR-----LLKIMNLQQIGRNYYNPNDPIDIPSHR 245
```

```
Db        136 ----------PQVLSSIMFLDLLLKKKPSETLFGFMHSFFTGENGVSLGGGVE----- 178

Qy        246 LVIWPGFTTSILQYENSIMLCTDVSHKVL-RSETVLDFM---------------------- 283

Db        179 --AWKGFYQSIRPNQGFMSVNVDISSSAFWRNDSLLQILMEYTDCSNVRDLTRFDLKRLS 236

Qy        284 --FNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKK---ADGS 338

Db        237 RKFRFLKVTCQH-------RNNVG----TDLANRVYSIEGFSSKSASDSFFVRRLNGEEQ 285

Qy        339 GVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLT---GLTDK 395

Db        286 KISVAEYFLENHNVRLQYPNLPCIL----------VKNGAMLPIEFCFVVKGQRYTAK 333

Qy        396 MRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKND-NVQRELRDWGLSFDSNLLSFSG 454

Db        334 LNSD-----QTANMIRFAVQRPFERVQQIDDFVHQMDWDTDPYLTQYGMKIQKKMLEVPA 388

Qy        455 RILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISV--KPLDNWLLI---YTRR-NYEAA 508

Db        389 RVLETPSIRYGGDCIE-RPVSGRWN--LRGKRFLDPPRAPIRSWAVMCFTSTRRLPMRGI 445

Qy        509 NSLIQNLFKVTPAMGMQ--MRKAIMIEVDDR--TEAYLRVLQQKV----TADTQIVVCLL 560

Db        446 ENFLQTYVQTLTSLGINFVMKKPPVLYADIRGSVEELCITLYKKAEQVGNAPPDYLFFIL 505

Qy        561 SSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGG---ELW 617

Db        506 DKNSPEPYGSIKRVCNTMLGVPSQCAISKHI--LQSKPQYCANLGMKINVKVGGINCSLI 563

Qy        618 RVDIPLKLV--MIVGIDCYH-DMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDG 674

Db        564 PKSNPLGNVPTLILGGDVYHPGVGATGVSIASIVASVDLNGCKYTAVSRSQPRHQEVIEG 623

Qy        675 LKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPR 734

Db        624 MKDIVVYLLQGFRAMTKQQPQRIIYFRDGTSEGQFLSVINDELSQIKEACHSLSPKYNPK 683

Qy        735 XTVIVVKKRVNTRFFAQS---GGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSP 791

Db        684 ILVCTTQKRHHARFFIKNKSDGDRNGNPLPGTIIEKHVTHPYQYDFYLISHPSLQGVSVP 743

Qy        792 THYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFL---------- 841

Db        744 VHYTVLHDEIQMPPDQFQTLCYNLCYVYARATSAVSLVPPVYYAHLVSNLARYQDVTADD 803

Qy        842 ---------VGQSIHREPNLSLSNRL 858

Db        804 TFVETSEASMDQEV--KPLLALSSKL 827
```

RESULT 4
H86438
protein T19E23.7 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: H86438
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H86438
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1014 <STO>
A;Cross-references: GB:AE005172; NID:g6692120; PIDN:AAF24585.1; GSPDB:GN00141

C;Genetics:
A;Gene: T19E23.7
A;Map position: 1

Query Match 11.8%; Score 533; DB 2; Length 1014;
Best Local Similarity 23.8%; Pred. No. 2.8e-31;
Matches 234; Conservative 161; Mismatches 398; Indels 192; Gaps 45;

```
Qy          3 GRARARA---RGRARGQETAQLVGSTASQQPG--------YIQPRPQ--PPPAEGELFGR 49

Db         44 GRGSERGGGNRGQGRGEQQDFRSQSQRGPPPGHGGRGTTQFQQPRPQVAPQPSQAPASYA 103

Qy         50 GRQRGTAGGTAKSQGLQI-------SAGFQELSXAERGGRRRDFH-DLGVNTRQNLDHVK 101

Db        104 GSVGGVAGRGAWGRKPQVPSDSASPSTSTTVVSEPVRVAEVMNLKPSVQVATSDRKEPMK 163

Qy        102 ESKTGSSGII----VRLSTNHFRLTSRPQWALYQYHIDYN---PLMEARRLRSALLFQH- 153

Db        164 RPDRG--GVVAVRRVNLYVNHYKVNFNPESVIRHYDVEIKGEIPTKKVSRFELAMVRDKV 221

Qy        154 -----EDLIGKCHAFDG-----TILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPP 203

Db        222 FTDNPDEFPLAMTAYDGQKNIFSAVELPTG-SYKVEYPKTEEMRGRSYTFTIKQVN---- 276

Qy        204 TSPTCLQFYNIIFRRLLKIMNLQQ--IGRNYYNPNDPI--------DIPS---------- 243

Db        277 --------------VLKLGDLKEYMTGRSSFNPRDVLQGMDVVMKEHPSKCMITVGKSF 321

Qy        244 ------------HRLVIWPGFTTSILQYENSIMLCTDVSHKVLR-SETVLDFMFNFYHQT 290

Db        322 FTRETEPDEDFRFGVIAAKGYRHTLKPTAQGLSLCLDYSVLAFRKAMSVIEYLKLYFNWS 381

Qy        291 EEHKFQEQ-VSKEXIGLVV--------------LTKYNIKTYRVDDIDWDQN--PKSTF 332

Db        382 DMRQFRRRDVEEELIGLKVTVNHRKNKQKLTIVGLSMQNTKDIKFDLIDQEGNEPPRKT- 440

Qy        333 KKADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIP-ELCYLTG 391

Db        441 --------SIVEYFRIKYGRHIVHKDIPCLDLGKNGRQN----------FVPMEFCDL-- 480

Qy        392 LTDKMRNDFNVMKDLAVHTR----LTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDS 447

Db        481 VEGQIYPKDNLDKDSALWLKKLSLVNPQQRQRNIDKMIK-ARNGPSGGEIIGNFGLKVDT 539

Qy        448 NLLSFSGRILQTE--KIHQGGKTFDYNP---QFADWSKETRGAPLISVKPLDNWLLI--- 499

Db        540 NMTPVEGRVLKAPSLKLAERGRVVREEPNPRQNNQWNLMKKGVTRGSI--VKHWAVLDFT 597

Qy        500 YTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVD--------DRTEAYLRVLQQKVT- 550

Db        598 ASERFNKMPNDFVDNLIDRCWRLGMQMEAPIVYKTSRMETLSNGNAIEELLRSVIDEASR 657

Qy        551 ----ADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVA--RTLGKQQTVMAIATKI 604

Db        658 KHGGARPTLVLCAM-SRKDDGYKTLKWIAETKLGLVTQCFLTGPATKGGDQ----YRANL 712

Qy        605 ALQMNCKMGG---ELWRVDIPLKL---VMIVGIDCYHDMTAGRR--SIAGFVASIN-EGM 655

Db        713 ALKMNAKVGGSNVELMDTFSFFKKEDEVMFIGADVNHPAARDKMSPSIVAVVGTLNWPEA 772

Qy        656 TRWFSRCIFQDRGQELVDGL-KVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVN 714

Db        773 NRYAARVIAQPHRKEEIQGFGDACLE-LVKAHVQATGKRPNKIVIFRDGVSDAQFDMVLN 831

Qy        715 YEVPQFLDCLKSIGR-GYNPRXTVIVVKKRVNTRFFA---QSGGRLQNPLPGTVIDVEVT 770

Db        832 VEL---LDVKLTFEKNGYNPKITVIVAQKRHQTRFFPATNNDGSDKGNVPSGTVVDTKVI 888

Qy        771 RPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPA 830

Db        889 HPYEYDFYLCSHHGGIGTSKPTHYYTLWDELGFTSDQVQKLIFEMCFTFTRCTKPVSLVP 948

Qy        831 PCQYAHKLAFLVGQSIH----REPN 851
```

```
Db        949 PVYYADMVAFR-GRMYHEASSREKN 972
```

RESULT 5
T52134
Zwille protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Nov-2000
C;Accession: T52134
R;Laux, T.
submitted to the EMBL Data Library, January 1998
A;Reference number: Z25975
A;Accession: T52134
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-988 <LAU>
A;Cross-references: EMBL:AJ223508; PIDN:CAA11429.1
C;Genetics:
A;Gene: zwille
A;Map position: 5
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 11.2%; Score 510; DB 2; Length 988;
Best Local Similarity 23.1%; Pred. No. 1.4e-29;
Matches 220; Conservative 147; Mismatches 422; Indels 164; Gaps 38;

```
Qy          4 RARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRGTAGGTAKSQ 63
Db         65 RSRRRNRG---GRKSDQ---GDVCMRPS-SRPRKPPPPSQ--------------TTSSA 102

Qy         64 GLQISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTS 123
Db        103 VSVATAG--EIVAVN--------HQMQMGVRKNSNFAPRPGFGTLGTKCIVKANHF-LAD 151

Qy        124 RPQWALYQYHIDYNPLMEARRLRSAL------LFQHEDLIGKCHAFD--------GTILF 169
Db        152 LPTKDLNQYDVTITPEVSSKSVNRAIIAELVRLYKESDLGRRLPAYDGRKSLYTAGELPF 211

Qy        170 LPKRLQQKVTE----VFSKTRNGEDVRITIT-------------LTNELPPTSPTCLQFY 212
Db        212 TWKEFSVKIVDEDDGIINGPKRERSYKVAIKFVARANMHHLGEFLAGKRADCPQEAVQIL 271

Qy        213 NIIFRRLLKIMNLQQIGRNYYNPNDPIDIP---SHRLVIWPGFTTSILQYENSIMLCTDV 269
Db        272 DIVLRE-LSVKRFCPVGRSFFSPD--IKTPQRLGEGLESWCGFYQSIRPTQMGLSLNIDM 328

Qy        270 -SHKVLRSETVLDFMFNFYHQ-------TEEHKFQEQVSKEXIGLVVLTKYNI-KTYRVD 320
Db        329 ASAAFIEPLPVIEFVAQLLGKDVLSKPLSDSDRVKIKKGLRGVKVEVTHRANVRRKYRVA 388

Qy        321 DIDWDQNPKSTFKKADGSGV-SFLEYYRKQYNQEITDLKQPVL-VSQPKRRRGPGGTLPG 378
Db        389 GLTTQPTRELMFPVDENCTMKSVIEYFQEMYGFTIQHTHLPCLQVGNQKK----ASYLPM 444

Qy        379 PAMLIPE-LCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYI----HKNDN 433
Db        445 EACKIVEGQRYTKRLNEKQ------------ITALLKVTCQRAEGQRNDILRTVQHNAYD 492

Qy        434 VQRELRDWGLSFDSNLLSFSGRILQTE--KIHQGGKTFDYNPQFADWSKETRGAPLISVK 491
Db        493 QDPYAKEFGMNISEKLASVEARILPAPWLKYHENGKEKDCLPQVGQWNMMNK--KMINGM 550

Qy        492 PLDNWLLIYTRRNYE--AANSLIQNLFKVTPAMGMQMRKAIMIEV----DDRTE-----A 540
Db        551 TVSRWACVNFSRSVQENVARGFCNELGQMCEVSGMEFNPEPVIPIYSARPDQVEKALKHV 610

Qy        541 YLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVAR---TLGKQQTV 597
Db        611 YHTSMNKTKGKELELLLAILPDNNGSLYGDLKRICETELGLISQCCLTKHVFKISKQ--- 667

Qy        598 MAIATKIALQMNCKMGGE--------LWRV----DIPLKLVMIVGIDCYHDMTA--GRRS 643
Db        668 --YLADVSLKINVKMGGRNTVLVDAISCRIPLVSDIP---TIIFGADVTHPENGEESSPS 722

Qy        644 IAGFVASIN-EGMTRWFSRCIFQDRGQELVDGLKVCLQAALR----------AWNSC 689
Db        723 IAAVVASQDWPEVTKYAGLVCAQAHRQELIQDLYKTWQDPVRGTVSGGMIRDLLISFRKA 782

Qy        690 NEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFF 749
Db        783 TGQKPLRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEPNYQPPVTFIVVQKRHHTRLF 842

Qy        750 A------QSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGL 803
Db        843 ANNHRDKNSTDRSGNILPGTVVDTKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDENNF 902

Qy        804 KPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSN 856
Db        903 TADGIQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRA--RFYLEPEIMQDN 953

RESULT 6
A84678
Argonaute (AGO1)-like protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: A84678
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: A84678
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-997 <STO>
A;Cross-references: GB:AE002093; NID:g4510428; PIDN:AAD21514.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g27880
A;Map position: 2
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 11.1%; Score 504.5; DB 2; Length 997;
Best Local Similarity 22.7%; Pred. No. 3.7e-29;
Matches 219; Conservative 153; Mismatches 407; Indels 186; Gaps 36;

```
Qy 3 GRARARARGR---ARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRGTAGG- 58
Db 46 GGSVSAGRGRGNVGRGENTGDL---TATQVP------------VASAVSGGRGRGNIGDP 90

Qy 59 ----TAKSQGLQI-SAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKT-------- 105
Db 91 TFSVASSSKTVSVASSSKEESKNTEVSETMSNLQITSTETKPEMTSLPPASSKAVTFPVR 150

Qy 106 ---GSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEA----RRLRSALLFQHED--L 156
Db 151 PGRGTLGKKVMVRANHF-LVQVADRDLYHYDVSINPEVISKTVNRNVMKLLVKNYKDSHL 209

Qy 157 IGKCHAFD--------GTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTC 208
Db 210 GGKSPAYDGRKSLYTAGPLPFDSKEFVVNLAEKRADGSSGKDRPFKVAVKN---VTSTDL 266

Qy 209 LQFYNIIFRR-------LLKIMNL----------QQIGRNYYNPNDPIDIPSHR------ 245
Db 267 YQLQQFLDRKQREAPYDTIQVLDVVLRDKPSNDYVSVGRSFFHTSLGKDARDGRGELGDG 326

Qy 246 LVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVL-DFMFNFYHQTEEHKFQEQVSKEXI 304
Db 327 IEYWRGYFQSLRLTQMGLSLNIDVSARSFYEPIVVTDFISKFLNIRDLNRPLRDSDR--- 383

Qy 305 GLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFL----------------EYY 346
Db 384 ---LKVKKVLRTLKVKLLHW-----NCTKSAKISGISSLPIRELRFTLEDKSEKTVVQYF 435
```
```

```
Qy        347 RKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELC-------YLTGLTDKMRND 399
Db        436 AEKYNYRVKYQALPAIQT---------GSDTRPVYLPMELCQIDEGQRYTKRLNEKQ--- 483

Qy        400 FNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQT 459
Db        484 ---VTALLKATCQRPPDRENSIKNLVVKNNYNDDLSKE---FGMSVTTQLASIEARVLPP 537

Qy        460 E--KIHQGGKTFDYNPQFADWS----KETRGAPLISVKPLDNWLLI--YTRRNYEAANSL 511
Db        538 PMLKYHDSGKEKMVNPRLGQWNMIDKKMVNGAKVTS------WTCVSFSTRIDRGLPQEF 591

Qy        512 IQNLFKVTPAMGMQMRKA-----IMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKD 566
Db        592 CKQLIGMCVSKGMEFKPQPAIPFISCPPEHIEEALLDI--HKRAPGLQLLIVIL-PDVTG 648

Qy        567 KYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGG------ELWRVD 620
Db        649 SYGKIKRICETELGIVSQCCQPRQVNKLN--KQYMENVALKINVKTGGRNTVLNDAIRRN 706

Qy        621 IPL---KLVMIVGIDCYHDMTA--GRRSIAGFVASIN-EGMTRWFSRCIFQDRGQELVDG 674
Db        707 IPLITDRPTIIMGADVTHPQPGEDSSPSIAAVVASMDWPEINKYRGLVSAQAHREEIIQD 766

Qy        675 LKVCLQAALR-------------AWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFL 721
Db        767 LYKLVQDPQRGLVHSGLIREHFIAFRRATGQIPQRIIFYRDGVSEGQFSQVLLHEMTAIR 826

Qy        722 DCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGG------RLQNPLPGTVIDVEVTRPEWY 775
Db        827 KACNSLQENYVPRVTFVIVQKRHHTRLFPEQHGNRDMTDKSGNIQPGTVVDTKICHPNEF 886

Qy        776 DFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYA 835
Db        887 DFYLNSHAGIQGTSRPAHYHVLLDENGFTADQLQMLTNNLCYTYARCTKSVSIVPPAYYA 946

Qy        836 HKLAF 840
Db        947 HLAAF 951
```

RESULT 7
T32079
hypothetical protein T07D3.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C;Accession: T32079
R;Lamar, B.; Wamsley, P.; Twyman, B.
submitted to the EMBL Data Library, July 1997
A;Description: The sequence of C. elegans cosmid T07D3.
A;Reference number: Z21121
A;Accession: T32079
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-891 <LAM>
A;Cross-references: EMBL:AF016682; PIDN:AAB66187.1; GSPDB:GN00020; CESP:T07D3.7
A;Experimental source: strain Bristol N2; clone T07D3
C;Genetics:
A;Gene: CESP:T07D3.7
A;Map position: 2
A;Introns: 34/2; 92/2; 206/1; 560/1; 879/3
C;Superfamily: rabbit translation initiation factor eIF-2C

 Query Match 11.0%; Score 500.5; DB 2; Length 891;
 Best Local Similarity 23.7%; Pred. No. 6.2e-29;
 Matches 215; Conservative 139; Mismatches 361; Indels 193; Gaps 36;

```
Qy         57 GGTAKSQGLQISAGFQELSXAERGGRRRDF----HDLGVN----TRQNLDHVKESKTGSS 108
Db          3 GDSLTSSSFMPDGGAETSSSSQLGGSAHGAIGTKPDAGVQFQCPVRPN--------HGVE 54

Qy        109 GIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTIL 168
Db         55 GRSILLRANHFAVRI-PGGSVQHYQIDVFPDKCPRRVNREVI----------GCLI 99

Qy        169 FLPKRLQQKVTEVFSKTRN----------GEDVRITITLTNE----------LPPTSPTC 208
Db        100 SSFSKYFTNIRPVYDGKRNMYTREPLPIGTEPMNFEVTLPGDSAVERKFSVTMKWIGQVC 159

Qy        209 L------------------QFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRL--- 246
Db        160 LSALDDAMEGRVRQVPHEAVQSIDVILRH-LPSLKYTPVGRSFFTPPGVMK-PGMQMHQE 217

Qy        247 -------VIWPGFTTSILQYENSIMLCTDVSHKVL-RSETVLDFMFNFYH---------- 288
Db        218 SKLGGGREVWFGFHQSVRPSQWKMMLNIDVSATAFYRAMPVIEFVAEVLELPVQALAERR 277

Qy        289 ---QTEEHKFQEQVSKEXIGL-VVLTKYNI--KTYRVDDIDWDQNPKSTFKKADGSGVS- 341
Db        278 ALSDAQRVKF----TKEIRGLKIEITHCGAVRRKYRVCNVTRRPAQTQTFPLQLETGQTI 333

Qy        342 -------FLEYYRKQYNQEITDLKQPVL----VSQPKRRRGPGGTLPGPAMLIPELC--- 387
Db        334 ECTVAKYFFDKYRIQ-------LKYPHLPCLQVGQEQKH----------TYLPPEVCDIV 376

Qy        388 ----YLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQREL----- 438
Db        377 PGQRCLKKLTDVQTSTM-----IKATARSAPE-REREICKLVS--------KAELSADPF 422

Qy        439 -RDWGLSFDSNLLSFSGRILQTEKIHQGG--KTFDYNPQFADWSKETRGAPLISVKPLDN 495
Db        423 AHEFGITINPAMTEVKGRVLSAPKLLYGGRHRATTALPNQGVW--DMRGKQFHTGMEVRT 480

Qy        496 W-LLIYTRRNYEAANSL---IQNLFKVTPAMGMQM-------RKAIMIEVDDRTEAYLRV 544
Db        481 WAIACFAQQSHVKENDLRMFTTQLQRISTDAGMPIIGTPMFCKYASGVE---QVEPMFKY 537

Qy        545 LQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKI 604
Db        538 LKQTYSAIQLIVVVL--PGKTPIYAEVKRVGDTVLGIATQCVQAK--NAIRTTPQTLSNL 593

Qy        605 ALQMNCKMGG------ELWRVDIPLKLVMIVGIDCYHDMTAGRR--SIAGFVASINEGMT 656
Db        594 CLKMNVKLGGVNSILLPNVRPRIFNEPVIFLGCDITHPAAGDTRKPSIAAVVGSMDAHPS 653

Qy        657 RWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYE 716
Db        654 RYAATVRVQQHRQEIITDLTYMVRELLVQFYRNTRFKPARIVVYRDGVSEGQLFNVLQYE 713

Qy        717 VPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFF----AQSGGRLQNPLPGTVIDVEVTRP 772
Db        714 LRAIREACVMLESGYQPGITFIAVQKRHHTRLFAADKADQVGKAFNIPPGTTVDVGITHP 773

Qy        773 EWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPC 832
Db        774 TEFDFFLCSHAGIQGTSRPSHYHVLWDDNDLTADELQQLTYQMCHTYVRCTRSVSIPAPA 833

Qy        833 QYAHKLAF 840
Db        834 YYAHLVAF 841
```

RESULT 8
T22391
hypothetical protein F48F7.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T22391; T23234
R;Coles, L.
submitted to the EMBL Data Library, February 1996
A;Reference number: Z19559
A;Accession: T22391
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1000 <WIL>
A;Cross-references: EMBL:Z69661; PIDN:CAA93496.1; GSPDB:GN00028; CESP:F48F7.1

```
A;Experimental source: clone F48F7
R;Coles, L.
submitted to the EMBL Data Library, February 1996
A;Reference number: Z19712
A;Accession: T23234
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1000 <WI2>
A;Cross-references: EMBL:Z69663; PIDN:CAA93512.1; GSPDB:GN00028; CESP:F48F7.1
A;Experimental source: clone K02B9
C;Genetics:
A;Gene: CESP:F48F7.1
A;Map position: X
A;Introns: 70/3; 128/2; 185/2; 673/1
C;Superfamily: rabbit translation initiation factor eIF-2C

  Query Match            10.9%;  Score 492;  DB 2;  Length 1000;
  Best Local Similarity  22.9%;  Pred. No. 3.2e-28;
  Matches  227;  Conservative  142;  Mismatches  361;  Indels  262;  Gaps   41;

Qy         18 TAQLVGSTASQQPGYIQPRPQPPPAEGELFG--RGRQRGT------AGGTAKSQGL---- 65
Db         56 SAQAAASTALQND--LEEIFNSPPTQPQTFSDVPQRQAGSLAPGVPIGNTSVSIGEPANT 113

Qy         66 ----------QISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRL 114
Db        114 LGGGLPGGAPGQLPGGNQ--SGIQFQCPRRPNH-----------------GVEGRSILL 153

Qy        115 STNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRL 174
Db        154 RANHFAVRI-PGGTIQHYQVDVTPDKCPRRVNREII--------------SCLISAFSKY 198

Qy        175 QQKVTEVFSKTRN---------GEDVRITITLTNELPPTSPTCLQF-------------- 211
Db        199 FTNIRPVYDGKRNMYTREPLPIGRE-RMDFDVT--LPGDSAVERQFSVSLKWVGQVSLST 255

Qy        212 -------------------YNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRLV----- 247
Db        256 LEDAMEGRVRQVPFEAVQAMDVILRH-LPSLKYTPVGRSFFSP--PVPNASGVMAGSCPP 312

Qy        248 --------------------------IWPGFTTSILQYENSIMLCTDVSHKVL-RSETV 279
Db        313 QASGAVAGGAHSAGQYHAESKLGGGREVWFGFHQSVRPSQWKMMLNIDVSATAFYRSMPV 372

Qy        280 LDFMFNFYH------------QTEEHKFQEQVSKEXIGLVVLTKYNI-------KTYRV 319
Db        373 IEFIAEVLELPVQALAERRALSDAQRVKF----TKEIRGL----KIEITHCGQMRRKYRV 424

Qy        320 DDIDWDQNPKSTFKKADGSGVS--------FLEYYRKQYNQEITDLKQPVL----VSQPK 367
Db        425 CNVTRRPAQTQTFPLQLETGQTIECTVAKYFYDKYRIQ-------LKYPHLPCLQVGQEQ 477

Qy        368 RRRGPGGTLPGPAMLIPELCYLT-------GLTDKMRNDFNVMKDLAVHTRLTPEQRQRE 420
Db        478 KH----------TYLPPEVCNIVPGQRCIKKLTDVQTSTM-----IKATARSAPE-RERE 521

Qy        421 VGRLI--------DYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKT-FDY 471
Db        522 ISNLVRKAEFSADPFAH----------EFGITINPAMTEVKGRVLSAPKLLYGGRTRATA 571

Qy        472 NPQFADWSKETRGAPLISVKPLDNW-LLIYTRRNYEAANSL---IQNLFKVTPAMGMQM- 526
Db        572 LPNQGVW--DMRGKQFHTGIDVRVWAIACFAQQQHVKENDLRMFTNQLQRISNDAGMPIV 629

Qy        527 ------RKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCP 580
Db        630 GNPCFCKYAVGVE---QVEPMFKYLKQNYSG-IQLVVVVL-PGKTPVYAEVKRVGDTVLG 684

Qy        581 TPSQCVVARTLGKQQTVMAIATKIALQMNCKMGG------ELWRVDIPLKLVMIVGIDCY 634
Db        685 IATQCVQAK--NAIRTTPQTLSNLCLKMNVKLGGVNSILLPNVRPRIFNEPVIFFGCDIT 742

Qy        635 HDMTAGRR--SIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEY 692
Db        743 HPPAGDSRKPSIAAVVGSMDAHPSRYAATVRVQQHRQEIISDLTYMVRELLVQFYRNTRF 802

Qy        693 MPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFA-- 750
Db        803 KPARIVVYRDGVSEGQFFNVLQYELRAIREACMMLERGYQPGITFIAVQKRHHTRLFAVD 862

Qy        751 --QSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHI 808
Db        863 KKDQVGKAYNIPPGTTVDVGITHPTEFDFYLCSHAGIQGTSRPSHYHVLWDDNNLTADEL 922

Qy        809 QRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAF 840
Db        923 QQLTYQMCHTYVRCTRSVSIPAPAYYAHLVAF 954


RESULT 9
A84668
Argonaute (AGO1)-like protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: A84668
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: A84668
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-930 <STO>
A;Cross-references: GB:AE002093; NID:g3885334; PIDN:AAC77862.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g27040
A;Map position: 2
C;Superfamily: rabbit translation initiation factor eIF-2C

  Query Match            10.5%;  Score 476.5;  DB 2;  Length 930;
  Best Local Similarity  21.2%;  Pred. No. 4.2e-27;
  Matches  184;  Conservative  170;  Mismatches  341;  Indels  173;  Gaps   34;

Qy        106 GSSGIIVRLSTNHFRL-TSRPQWALYQYHI-----DYNPLMEARRLRSALLFQ-----HE 154
Db         63 GTRGQKIPLLTNHFKVDVANLQGHFFHYSVALFYDDGRP-VEQKGVGRKILDKVHQTYHS 121

Qy        155 DLIGKCHAFDG-----TILFLPKRLQ--QKVTEVFSKTR-------------NG------ 188
Db        122 DLDGKEFAYDGEKTLFTYGALPSNKMDFSVVLEEVSATSKDFVSRANGNGSPNGNESPSD 181

Qy        189 ------------EDVRITITLTNELP-------------PTSPTCLQFYNIIFRRLLKIM 223
Db        182 GDRKRLRRPNRSKNFRVEISYAAKIPLQALANAMRGQESENSQEAIRVLDIILRQHAARQ 241

Qy        224 NLQQIGRNYYNPNDPIDIP--SHRLVIWPGFTTSILQYENSIMLCTDV-SHKVLRSETVL 280
Db        242 GCLLVRQSFFH-NDPTNCEPVGGNILGCRGFHSSFRTTQGGMSLNMDVTTTMIIKPGPVV 300

Qy        281 DFMFNFYHQTEEHKFQEQVSKEXI-GLVVLTKYNIKTYRVDDIDWDQNPKSTFK------ 333
Db        301 DFLIANQNARDPYSIDWSKAKRTLKNLRVKVSPSGQEFKITGLSDKPCREQTFELKKRNP 360

Qy        334 ----KADGSGVSFLEYYRK------QYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLI 383
Db        361 NENGEFETTEVTVADYFRDTRHIDLQYSADLPCIN----VGKPKR----------PTYIP 406

Qy        384 PELC-------YLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQR 436
Db        407 LELCALVPLQRYTKALTTFQRS------ALVEKSRQKPQERMTVLSKALKV--SNYDAEP 458

Qy        437 ELRDWGLSFDSNLLSFSGRILQTEKIHQ--GGKTFDYNPQFADWSKETRGAPLISVKPLD 494
Db        459 LLRSCGISISSNFTQVEGRVLPAPKLKMGCGSETFPRNGRWNFNNKE-----FVEPTKIQ 513
```

```
Qy        495 NWLLIYTRRNYEA---ANSLIQNLFKVTPAMGMQMRKAIMIEVDD----------RTEAY 541
Db        514 RWVVV----NFSARCNVRQVVDDLIKIGGSKGIEIASPFQVFEEGNQFRRAPPMIRVENM 569

Qy        542 LRVLQQKVTADTQIVVCLLSSNR-KDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAI 600
Db        570 FKDIQSKLPGVPQFILCVLPDKKNSDLYGPWKKKNLTEFGIVTQCMAPTRQPNDQYL--- 626

Qy        601 ATKIALQMNCKMGG--ELWRVD----------IPLKLVMIVGIDCYHDMTAGRRSIAGFV 648
Db        627 -TNLLLKINAKLGGLNSMLSVERTPAFTVISKVP---TIILGMDVSHG-SPGQSDVPSIA 681

Qy        649 ASINEG----MTRWFSRCIFQDRGQELVDGL----------KVCLQAALRAWNSCNEYMPS 695
Db        682 AVVSSREWPLISKYRASVRTQPSKAEMIESLVKKNGTEDDGIIKELLVDFYTSSNKRKPE 741

Qy        696 RIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGR 755
Db        742 HIIIFRDGVSESQFNQVLNIELDQIIEACKLLDANWNPKFLLLVAQKNHHTKFFQPTSP- 800

Qy        756 LQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKL 815
Db        801 -ENVPPGTIIDNKICHPKNNDFYLCAHAGMIGTTRPTHYHVLYDEIGFSADELQELVHSL 859

Qy        816 CHIYYNWPGVIRVPAPCQYAHKLAFLVG 843
Db        860 SYVYQRSTSAISVVAPICYAHLAAAQLG 887
```

RESULT 10
JC6569
translation initiation factor eIF-2C - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 21-Jul-2000
C;Accession: PC6505; JC6569
R;Zou, C.; Zhang, Z.; Wu, S.; Osterman, J.C.
Gene 211, 187-194, 1998
A;Title: Molecular cloning and characterization of a rabbit eIF2C protein.
A;Reference number: JC6569; MUID:98267198; PMID:9602122
A;Accession: PC6505
A;Molecule type: protein
A;Residues: 336-359;380-409;694-711 <ZOU1>
A;Experimental source: liver
A;Accession: JC6569
A;Molecule type: mRNA
A;Residues: 1-813 <ZOU2>
A;Cross-references: GB:AF005355; NID:g3253158; PIDN:AAC24323.1; PID:g3253159
C;Superfamily: rabbit translation initiation factor eIF-2C
C;Keywords: liver

 Query Match 10.5%; Score 474; DB 2; Length 813;
 Best Local Similarity 23.9%; Pred. No. 5.2e-27;
 Matches 191; Conservative 145; Mismatches 343; Indels 120; Gaps 32;

```
Qy        125 PQWALYQYHIDYNPLMEARRLRSAL---LFQH--EDLIG-KCHAFDG-----TILFLPKR 173
Db          4 PKIDIYHYELDIKPEKCPRRVNREIVEHMVQHFKAQIFGDRKPVFDGRKNLYTAMPLPIG 63

Qy        174 LQQKVTEVFSKTRNGED--VRITI------------TLTNELPPTSPTCLQFYNIIFRR 218
Db         64 -REKVELEVTLPGEGKDRIFKVSIKWVSCVSLQALHDALSGRLPSVPFETIQALDVVMRH 122

Qy        219 LLKIMNLQQIGRNYYNPNDPIDIP-SHRLVIWPGFTTSILQYENSIMLCTDVSHKVL-RS 276
Db        123 -LPSMRYTPVGRSFFTASEGCSNPLGGGREVWFGFHQSVRPSLWKMMLNIDVSATAFYKA 181

Qy        277 ETVLDF---MFNFYHQTEEHK-----FQEQVSKEXIGLVVLTKY---NIKTYRVDDIDWD 325
Db        182 QPVIEFVCEVLDFKSIEEQQKPLTDSQRVKFTKEIKGLKVEITHCGQMKRKYRVCNVTRR 241

Qy        326 QNPKSTFKKADGSG----VSFLEYYRKQYNQEITDLKQPVL-VSQPKRRRGPGGTLPGPA 380
Db        242 PASHQTFPLQQESGQTVECTVAQYFKDRHKLVLRYPHLPCLQVGQEQKH----------T 291

Qy        381 MLIPELCYLTG-------LTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDN 433
Db        292 YLPLEVCNIVAGQRCIKKLTD---NQTSTM--IRATARSAPD-RQEEISKLMR--SASFN 343

Qy        434 VQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPL 493
Db        344 TDPYVREFGIMVKDEMTDVTGRVLQPPSILYGGR-----------NKAIATPVQGVWDM 391

Qy        494 DN----------WLLIYTRRNYEAA----NSLIQNLFKVTPAMGMQMRK----AIMIEV 534
Db        392 RNKQFHTGIEIKVWAIACFAPQRQCTEVHLKSFTEQLRKISRDAGMPIQGQPCFCKYAQG 451

Qy        535 DDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQ 594
Db        452 ADSVGPMFRHL-KNTYAGLQLVVVIL-PGKTPVYAEVKRVGDTVLGMATQCVQMKNV--Q 507

Qy        595 QTVMAIATKIALQMNCKMGG------ELWRVDIPLKLVMIVGIDCYHDMTAG---RRSIA 645
Db        508 RTTPQTLSNLCLKINVKLGGVNNILLPQGRPPVFQQPVIFLGADVTHP-PAGDGKKPSIA 566

Qy        646 GFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVG 705
Db        567 AVVGSMDAHPNRYCATVRVQQHRQEIIQDLAAMVRELLIQFYKSTRFKPTRIIFYRDGVS 626

Qy        706 DGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFA----QSGGRLQNPLP 761
Db        627 EGQFQQVLHHELLAIREACIKLEKDYQPGITFIVVQKRHHTRLFCTDKNERVGKSGNIPA 686

Qy        762 GTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYN 821
Db        687 GTTVDTKITHPTEFDFYLCSHAGIQGTSRPSHYHVLWDDNRFSSDELQILTYQLCHTYVR 746

Qy        822 WPGVIRVPAPCQYAHKLAF 840
Db        747 CTRSVSIPAPAYYAHLVAF 765
```

RESULT 11
T23164
hypothetical protein T22B3.2a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T23164; T25099
R;Cottage, A.
submitted to the EMBL Data Library, January 1996
A;Reference number: Z19701
A;Accession: T23164
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1032 <WIL>
A;Cross-references: EMBL:Z68750; PIDN:CAA92969.1; GSPDB:GN00022; CESP:T22B3.2a
A;Experimental source: clone K01A6
R;Lennard, N.
submitted to the EMBL Data Library, December 1995
A;Reference number: Z19981
A;Accession: T25099
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1032 <WI2>
A;Cross-references: EMBL:Z68300; PIDN:CAA92618.1; GSPDB:GN00022; CESP:T22B3.2a
A;Experimental source: clone T22B3
C;Genetics:
A;Gene: CESP:T22B3.2a
A;Map position: 4
A;Introns: 19/1; 71/3; 129/2; 424/2; 496/3; 607/1; 666/1; 753/3; 795/2; 961/2; 1003/1
C;Superfamily: rabbit translation initiation factor eIF-2C

 Query Match 9.5%; Score 430.5; DB 2; Length 1032;
 Best Local Similarity 22.1%; Pred. No. 1.3e-23;
 Matches 203; Conservative 144; Mismatches 364; Indels 209; Gaps 37;

```
Qy        106 GSSGIIVRLSTNHFRL-TSRPQWALYQYHID-YNPLMEARRL---RSALLF-----QHED 155
Db         88 GTIGRKIPVKSNFFAVDLKNPKMVVVQYHVEVHHP--GCRKLDKDEMRIIFWKAVSDHPN 145

Qy        156 LIGK--CHAFDGT-ILFLPKRLQ--------QKVTEVFSKTRNGEDVRITI--------- 195
Db        146 IFHNKFALAYDGAHQLYTVARLEFPDDQGSVRLDCEATLPKDNRDRTRCAISIQNVGPVL 205

Qy        196 -----TLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYN----------PNDPI 239
Db        206 LEMQRTRTNNLDERVLTPIQILDIICRQSLTCPLLKN-SANFYTWKSSCYRIPTAAGQAL 264

Qy        240 DIPSHRLVIWPGFTTSILQYEN-SIMLCTDVSHKVLRSE--TVLDFMFNFYHQ------- 289
Db        265 DLEGGK-EMWTGFFSSAHIASNYRPLLNIDVAHTAFYKTRITVLQFMCDVLNERTSKPNR 323

Qy        290 ------------------------------------------------------TEEH 293
Db        324 NNPRGPGAPGGYRGGRGGARGGSYQNFGNRGPPGANVRDDFGGNGLTFTMDTLSRDTQLS 383

Qy        294 KFQEQVSKEXI-GLVVLTKYN---IKTYRVDDIDWDQNP--KSTFKKADGSG----VSFL 343
Db        384 SFETRIFGDSIRGMKIRATHRPNAIRVYKVNSL---QLPADKLMFQGIDEEGRQVVCSVA 440

Qy        344 EYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIP-ELC-------YLTGLTDK 395
Db        441 DYFSEKYG----PLKYPKL---PCLHVGP----PTRNIFLPMEHCLIDSPQKYNKKMTEK 489

Qy        396 MRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGR 455
Db        490 QTS--AIIKAAAVDA----TQREDRIKQLA--AQASFGTDPFLKEFGVAVSSQMIETSAR 541

Qy        456 ILQTEKIHQGGKTFDYN----PQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSL 511
Db        542 VIQPPPIMFGGNNRSINPVVFPKDGSWSMDHQTLYMPATCRSYSMIALVDPRDQTSLQTF 601

Qy        512 IQNLFKVTPAMGMQMRK---------------AIMIEVDDRTEAYLRVLQQKVTADT-QI 555
Db        602 CQSLTMKATAMGMNFPRWPDLVKYGRSKEDVCTLFTEIAD---------EYRVTNTVCDC 652

Qy        556 VVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGG- 614
Db        653 IIVVLQSKNSDIYMTVKEQSDIVHGIMSQCVLMKNVSR--PTPATCANIILKLNMKMGGI 710

Qy        615 ----------ELWRVDIPLKLVMIVGIDCYHDMTAGRR----SIAGFVASINEGMTRWFS 660
Db        711 NSRIVADQITNKYLVDQP---TMVVGIDVTHPTQAEMRMNMPSVAAIVANVDLLPQSYGA 767

Qy        661 RCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQF 720
Db        768 NVKVQKKCRESVVYLLDAIRERIITFYRHTKQKPARIIVYRDGVSEGQFSEVLREEIQSI 827

Qy        721 LDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQ----SGGRLQNPLPGTVIDVEVTRPEWYD 776
Db        828 RTACLAIAEDFRPPITYIVVQKRHHARIFCKFPNDMVGKAKNVPPGTTVDTGIVSPEGFD 887

Qy        777 FFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAH 836
Db        888 FYLCSHYGVQGTSRPARYHVLLDECKFTADEIQNITYGMCHTYGRCTRSVSIPTPVYYAD 947

Qy        837 KLAFLVGQSIHREPNLSLSN 856
Db        948 LVATRARCHIKRKLGLADNN 967
```

RESULT 12
T23165
hypothetical protein T22B3.2b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T23165; T25100
R;Cottage, A.
submitted to the EMBL Data Library, January 1996
A;Reference number: Z19701
A;Accession: T23165
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1035 <WIL>
A;Cross-references: EMBL:Z68750; PIDN:CAA92970.1; GSPDB:GN00022; CESP:T22B3.2b
A;Experimental source: clone K01A6
R;Lennard, N.
submitted to the EMBL Data Library, December 1995
A;Reference number: Z19981
A;Accession: T25100
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1035 <WI2>
A;Cross-references: EMBL:Z68300; PIDN:CAA92619.1; GSPDB:GN00022; CESP:T22B3.2b
A;Experimental source: clone T22B3
C;Genetics:
A;Gene: CESP:T22B3.2b
A;Map position: 4
A;Introns: 19/1; 74/3; 132/2; 427/2; 499/3; 610/1; 669/1; 756/3; 798/2; 964/2; 1006/1
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 9.5%; Score 430.5; DB 2; Length 1035;
Best Local Similarity 22.1%; Pred. No. 1.4e-23;
Matches 203; Conservative 144; Mismatches 364; Indels 209; Gaps 37;

```
Qy        106 GSSGIIVRLSTNHFRL-TSRPQWALYQYHID-YNPLMEARRL---RSALLF-----QHED 155
Db         91 GTIGRKIPVKSNFFAVDLKNPKMVVVQYHVEVHHP--GCRKLDKDEMRIIFWKAVSDHPN 148

Qy        156 LIGK--CHAFDGT-ILFLPKRLQ--------QKVTEVFSKTRNGEDVRITI--------- 195
Db        149 IFHNKFALAYDGAHQLYTVARLEFPDDQGSVRLDCEATLPKDNRDRTRCAISIQNVGPVL 208

Qy        196 -----TLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYN----------PNDPI 239
Db        209 LEMQRTRTNNLDERVLTPIQILDIICRQSLTCPLLKN-SANFYTWKSSCYRIPTAAGQAL 267

Qy        240 DIPSHRLVIWPGFTTSILQYEN-SIMLCTDVSHKVLRSE--TVLDFMFNFYHQ------- 289
Db        268 DLEGGK-EMWTGFFSSAHIASNYRPLLNIDVAHTAFYKTRITVLQFMCDVLNERTSKPNR 326

Qy        290 ------------------------------------------------------TEEH 293
Db        327 NNPRGPGAPGGYRGGRGGARGGSYQNFGNRGPPGANVRDDFGGNGLTFTMDTLSRDTQLS 386

Qy        294 KFQEQVSKEXI-GLVVLTKYN---IKTYRVDDIDWDQNP--KSTFKKADGSG----VSFL 343
Db        387 SFETRIFGDSIRGMKIRATHRPNAIRVYKVNSL---QLPADKLMFQGIDEEGRQVVCSVA 443

Qy        344 EYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIP-ELC-------YLTGLTDK 395
Db        444 DYFSEKYG----PLKYPKL---PCLHVGP----PTRNIFLPMEHCLIDSPQKYNKKMTEK 492

Qy        396 MRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGR 455
Db        493 QTS--AIIKAAAVDA----TQREDRIKQLA--AQASFGTDPFLKEFGVAVSSQMIETSAR 544

Qy        456 ILQTEKIHQGGKTFDYN----PQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSL 511
Db        545 VIQPPPIMFGGNNRSINPVVFPKDGSWSMDHQTLYMPATCRSYSMIALVDPRDQTSLQTF 604

Qy        512 IQNLFKVTPAMGMQMRK---------------AIMIEVDDRTEAYLRVLQQKVTADT-QI 555
Db        605 CQSLTMKATAMGMNFPRWPDLVKYGRSKEDVCTLFTEIAD---------EYRVTNTVCDC 655

Qy        556 VVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGG- 614
Db        656 IIVVLQSKNSDIYMTVKEQSDIVHGIMSQCVLMKNVSR--PTPATCANIILKLNMKMGGI 713

Qy        615 ----------ELWRVDIPLKLVMIVGIDCYHDMTAGRR----SIAGFVASINEGMTRWFS 660
Db        714 NSRIVADQITNKYLVDQP---TMVVGIDVTHPTQAEMRMNMPSVAAIVANVDLLPQSYGA 770
```

```
Qy        661 RCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQF 720
Db        771 NVKVQKKCRESVVYLLDAIRERIITFYRHTKQKPARIIVYRDGVSEGQFSEVLREEIQSI 830

Qy        721 LDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQ----SGGRLQNPLPGTVIDVEVTRPEWYD 776
Db        831 RTACLAIAEDFRPPITYIVVQKRHHARIFCKFPNDMVGKAKNVPPGTTVDTGIVSPEGFD 890

Qy        777 FFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAH 836
Db        891 FYLCSHYGVQGTSRPARYHVLLDECKFTADEIQNITYGMCHTYGRCTRSVSIPTPVYYAD 950

Qy        837 KLAFLVGQSIHREPNLSLSN 856
Db        951 LVATRARCHIKRKLGLADNN 970
```

RESULT 13
T01113
translation initiation factor eIF-2C homolog T21L14.12 - Arabidopsis thaliana
N;Alternate names: Argonaute (AGO1)-like protein
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 16-Feb-2001
C;Accession: T01113; D84739
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, December 1997
A;Description: Arabidopsis thaliana chromosome II BAC T21L14 genomic sequence.
A;Reference number: Z14209
A;Accession: T01113
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-887 <ROU>
A;Cross-references: EMBL:AC003033; NID:g2702261; PID:g2702284
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: D84739
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-887 <STO>
A;Cross-references: GB:AE002093; NID:g2702284; PIDN:AAB91987.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g32940; T21L14.12
A;Map position: 2
A;Introns: 61/3; 130/2; 203/2; 248/1; 281/3; 314/2; 367/3; 406/3; 442/3; 463/3; 501/3; 5
C;Superfamily: rabbit translation initiation factor eIF-2C

```
  Query Match             9.3%;  Score 422;  DB 2;  Length 887;
  Best Local Similarity   22.0%;  Pred. No. 4.6e-23;
  Matches  196;  Conservative  151;  Mismatches  354;  Indels  190;  Gaps   41;

Qy         88 DLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSR-PQWALYQYHI----DYNPLMEA 142
Db         23 DYDITTRRGV--------GTTGNPIELCTNHFNVSVRQPDVVFYQYTVSITTENGDAVDG 74

Qy        143 RRLRSALLFQ-----HEDLIGKCHAFDG-TILFLPKRLQQK-------VTEVFSKTRNG- 188
Db         75 TGISRKLMDQLFKTYSSDLDGKRLAYDGEKTLYTVGPLPQNEFDFLVIVEGSFSKRDCGV 134

Qy        189 ------------------EDVRITITLTNELP--------------PTSPTCLQFYNI 214
Db        135 SDGGSSSGTCKRSKRSFLPRSYKVQIHYAAEIPLKTVLGTQRGAYTPDKSAQDALRVLDI 194

Qy        215 IFRRLLKIMNLQQIGRN---------YYNPNDPIDIPSHRLVIWPGFTTSILQYENSIML 265
Db        195 VLR--------QQAAERGCLLVRQAFFHSDGHPMKVGGGVIGI-RGLHSSFRPTHGGLSL 245

Qy        266 CTDVS-HKVLRSETVLDFMFNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIK-TYRVDD-- 321
Db        246 NIDVSTTMILEPGPVIEFL----KANQSVETPRQIDWIKVAAKMLKHMRVKATHRNMEFK 301

Qy        322 -IDWDQNPKS----TFKKADGS------GVSFLEYYRKQYNQEITDLKQPVL-VSQPKRR 369
Db        302 IIGLSSKPCNQQLFSMKIKDGEREVPIREITVYDYFKQTYTEPISSAYFPCLDVGKPDR- 360

Qy        370 RGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTR--LTPEQRQREVGR---- 423
Db        361 ---------PNYLPLEFCNLVSLQ--------RYTKPLSGRQRVLLVESSRQKPLERIKTL 404

Qy        424 ---LIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSK 480
Db        405 NDAMHTYCYDKDPF---LAGCGISIEKEMTQVEGRVLKPPML-KFGKNEDFQPCNGRWNF 460

Qy        481 ETRGAPLISVKPLDNWLLIYTRRNYE---AANSLIQNLFKVTPAMGMQM-RKAIMIEVD- 535
Db        461 NNK--MLLEPRAIKSWAIV----NFSFPCDSSHISRELISCGMRKGIEIDRPFALVEEDP 514

Qy        536 --------DRTEAYLRVLQQKVTADTQIVVCLLSSNR-KDKYDAIKKYLCTDCPTPSQCV 586
Db        515 QYKKAGPVERVEKMIATMKLKFPDPPHFILCILPERKTSDIYGPWKKICLTEEGIHTQCI 574

Qy        587 VARTLGKQQTVMAIATKIALQMNCKMG------GELWRVDIPL--KL-VMIVGIDCYHDM 637
Db        575 CPIKISDQ-----YLTNVLLKINSKLGGINSLLGIEYSYNIPLINKIPTLILGMDVSHG- 628

Qy        638 TAGRRSIAGFVASINEG----MTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNS-CNEY 692
Db        629 PPGRADVPSVAAVVGSKCWPLISRYRAAVRTQSPRLEMIDSLFQPIENTEKGDNGIMNEL 688

Qy        693 M-----------PSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVK 741
Db        689 FVEFYRTSRARKPKQIIIFRDGVSESQFEQVLKIEVDQIIKAYQRLGESDVPKFTVIVAQ 748

Qy        742 KRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNS 801
Db        749 KNHHTKLFQAKGP--ENVPAGTVVDTKIVHPTNYDFYMCAHAGKIGTSRPAHYHVLLDEI 806

Qy        802 GLKPDHIQR----LTYKLCHIYYNWPGVIRV----PAPCQYAHKLAFLVGQ 844
Db        807 GFSPDDLQNLIHSLSYKLLNSIFNVSSLLCVFVLSVAPVRYAHLAAAQVAQ 857
```

RESULT 14
S41013
hypothetical protein ZK757.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Feb-2000
C;Accession: S41013
R;Thomas, K.
submitted to the EMBL Data Library, December 1993
A;Reference number: S41011
A;Accession: S41013
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-958 <THO>
A;Cross-references: EMBL:Z29121
C;Genetics:
A;Introns: 52/2; 350/2; 422/3; 592/1; 679/3; 721/2; 887/2; 929/1
C;Superfamily: rabbit translation initiation factor eIF-2C

```
  Query Match             9.3%;  Score 421;  DB 2;  Length 958;
  Best Local Similarity   21.6%;  Pred. No. 6.1e-23;
  Matches  199;  Conservative  148;  Mismatches  364;  Indels  212;  Gaps   37;

Qy        106 GSSGIIVRLSTNHFRL-TSRPQWALYQYHID-YNPLMEARRL---RSALLF-----QHED 155
Db         11 GTIGRQIPVKSNFFAMDLKNPKMVVIQYHVEIHHP--GCRKLDKDEMRIIFWKAVSDHPN 68

Qy        156 LIGK--CHAFDGT-ILFLPKRLQ--------QKVTEVFSKTRNGEDVRITI--------- 195
Db         69 IFHNKFALAYDGAHQLYTVARLEFPDDQGSVRLDCEASLPKDNRDRTRCAISIQNVGPVL 128
```

```
Qy        196 -----TLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYN----------PNDPI 239
                  | || |     | :|  :|| |: |    |:    |:|                :
Db        129 LEMQRTRTNNLDERVLTPIQILDIICRQSLTCPLLKN-SANFYTWKSSCYRIPTAAGQAL 187

Qy        240 DIPSHRLVIWPGFTTSILQYEN-SIMLCTDVSHKVLRSE--TVLDFMFNFYHQ------- 289
                :   :  :| || :|     |   :|  ||:|        ||| || :  ::
Db        188 DLEGGK-EMWTGFFSSAHIASNYRPLLNIDVAHTAFYKTRITVLQFMCDVLNERTSKPNR 246

Qy        290 ----------------------------------------------------------T 290
                                                                           |
Db        247 NNPRGPGGPGGPGGYRGGRGGGRGGSYGNFGNRGPPGANVRDDFGGNGLTFTMDTLSRDT 306

Qy        291 EEHKFQEQVSKEXI-GLVVLTKYN---IKTYRVDDIDWDQNP--KSTFKKADGSG----V 340
                :   |: ::  : | |: :    :    |: |:|: :   | |  |  |:  |  |
Db        307 QLSSFETRIFGDAIRGMKIRAAHRPNAIRVYKVNSL---QLPADKLMFQGIDEEGRQVVC 363

Qy        341 SFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIP-ELC-------YLTGL 392
                |  :|: ::|      || | |   |    ||    |   : :| | |       |   :
Db        364 SVADYFSEKYG----PLKYPKL---PCLHVGP----PTRNIFLPMEHCLIDSPQKYNKKM 412

Qy        393 TDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSF 452
                ::|  :   ::|  ||       ||:  : :|      :      |:::|::  | ::
Db        413 SEKQTS--AIIKAAAVDA----TQREDRIKQLA--AQASFGTDPFLKEFGVAVSSQMIQT 464

Qy        453 SGRILQTEKIHQGGKTFDYN----PQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA 508
                : |::|   |  ||    |    |:  |: : :   : :     : : :   |:  :
Db        465 TARVIQPPPIMFGGNNRSVNPVVFPKDGSWTMDNQTLYMPATCRSYSMIALVDPRDQTSL 524

Qy        509 NSLIQNLFKVTPAMGMQMRK--------------AIMIEVDDRTEAYLRVLQQKVTADT 553
                 :  |:|     ||||   :              :  |: |       : :||
Db        525 QTFCQSLTMKATAMGMNFPRWPDLVKYGRSKEDVCTLFTEIAD---------EYRVTNTV 575

Qy        554 -QIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKM 612
                 :: :| |   | |  :|:         ||||: : : :     |    | |::| ||
Db        576 CDCIIVVLQSKNSDIYMTVKEQSDIVHGIMSQCVLMKNVSR--PTPATCANIVLKLNMKM 633

Qy        613 GG-----------ELWRVDIPLKLVMIVGIDCYHDMTAGRR----SIAGFVASINEGMTR 657
                ||             : || |    |:|||| |   |  |    |:|  ||:::
Db        634 GGINSRIVADKITNKYLVDQP---TMVVGIDVTHPTQAEMRMNMPSVAAIVANVDLLPQS 690

Qy        658 WFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEV 717
                : :   | : :| |  |   ::  :   :    |:|||||||| :||   ::  |:
Db        691 YGANVKVQKKCRESVVYLLDAIRERIITFYRHTKQKPARIIVYRDGVSEGQFSEVLREEI 750

Qy        718 PQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQ----SGGRLQNPLPGTVIDVEVTRPE 773
                    :|   : |  | |||:|| : | | :      |: :|  ||| :|  :  ||
Db        751 QSIRTACLAIAEDFRPPITYIVVQKRHHARIFCKYQNDMVGKAKNVPPGTTVDTGIVSPE 810

Qy        774 WYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQ 833
                 :||:: |     |:  |  |:|: |      | || :|| :|| |      : :| |
Db        811 GFDFYLCSHYGVQGTSRPARYHVLLDECKFTADEIQSITYGMCHTYGRCTRSVSIPTPVY 870

Qy        834 YAHKLAFLVGQSIHREPNLSLSN 856
                ||  :|     : |:  |: :|
Db        871 YADLVATRARCHVKRKLGLADNN 893


RESULT 15
D88568
protein ZK757.3 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C;Accession: D88568
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: D88568
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1040 <STO>
A;Cross-references: GB:chr_III; PIDN:CAA82941.1; PID:g3877004; GSPDB:GN00021; CESP:ZK
C;Genetics:
A;Gene: ZK757.3
A;Map position: 3
C;Superfamily: rabbit translation initiation factor eIF-2C

  Query Match             9.3%;  Score 421;  DB 2;  Length 1040;
  Best Local Similarity   21.6%;  Pred. No. 7e-23;
  Matches  199;  Conservative  148;  Mismatches  364;  Indels  212;  Gaps   37;

Qy        106 GSSGIIVRLSTNHFRL-TSRPQWALYQYHID-YNPLMEARRL---RSALLF-----QHED 155
                |: |  : : :| | :    |:  : |||:: ::|    |:|      ::|      | :
Db         93 GTIGRQIPVKSNFFAMDLKNPKMVVIQYHVEIHHP--GCRKLDKDEMRIIFWKAVSDHPN 150

Qy        156 LIGK--CHAFDGT-ILFLPKRLQ--------QKVTEVFSKTRNGEDVRITI--------- 195
                :       |:||   |:   ||:       :   |      | :  | |
Db        151 IFHNKFALAYDGAHQLYTVARLEFPDDQGSVRLDCEASLPKDNRDRTRCAISIQNVGPVL 210

Qy        196 -----TLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYN----------PNDPI 239
                    | || |     | :|  :|| |: |    |:    |:|                :
Db        211 LEMQRTRTNNLDERVLTPIQILDIICRQSLTCPLLKN-SANFYTWKSSCYRIPTAAGQAL 269

Qy        240 DIPSHRLVIWPGFTTSILQYEN-SIMLCTDVSHKVLRSE--TVLDFMFNFYHQ------- 289
                |:   :  :| || :|     |   :|  ||:|        ||| || :  ::
Db        270 DLEGGK-EMWTGFFSSAHIASNYRPLLNIDVAHTAFYKTRITVLQFMCDVLNERTSKPNR 328

Qy        290 ----------------------------------------------------------T 290
                                                                           |
Db        329 NNPRGPGGPGGPGGYRGGRGGGRGGSYGNFGNRGPPGANVRDDFGGNGLTFTMDTLSRDT 388

Qy        291 EEHKFQEQVSKEXI-GLVVLTKYN---IKTYRVDDIDWDQNP--KSTFKKADGSG----V 340
                :   |: ::  : | |: :    :    |: |:|: :   | |  |  |:  |  |
Db        389 QLSSFETRIFGDAIRGMKIRAAHRPNAIRVYKVNSL---QLPADKLMFQGIDEEGRQVVC 445

Qy        341 SFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIP-ELC-------YLTGL 392
                |  :|: ::|      || | |   |    ||    |   : :| | |       |   :
Db        446 SVADYFSEKYG----PLKYPKL---PCLHVGP----PTRNIFLPMEHCLIDSPQKYNKKM 494

Qy        393 TDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSF 452
                ::|  :   ::|  ||       ||:  : :|      :      |:::|::  | ::
Db        495 SEKQTS--AIIKAAAVDA----TQREDRIKQLA--AQASFGTDPFLKEFGVAVSSQMIQT 546

Qy        453 SGRILQTEKIHQGGKTFDYN----PQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA 508
                : |::|   |  ||    |    |:  |: : :   : :     : : :   |:  :
Db        547 TARVIQPPPIMFGGNNRSVNPVVFPKDGSWTMDNQTLYMPATCRSYSMIALVDPRDQTSL 606

Qy        509 NSLIQNLFKVTPAMGMQMRK--------------AIMIEVDDRTEAYLRVLQQKVTADT 553
                 :  |:|     ||||   :              :  |: |       : :||
Db        607 QTFCQSLTMKATAMGMNFPRWPDLVKYGRSKEDVCTLFTEIAD---------EYRVTNTV 657

Qy        554 -QIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKM 612
                 :: :| |   | |  :|:         ||||: : : :     |    | |::| ||
Db        658 CDCIIVVLQSKNSDIYMTVKEQSDIVHGIMSQCVLMKNVSR--PTPATCANIVLKLNMKM 715

Qy        613 GG-----------ELWRVDIPLKLVMIVGIDCYHDMTAGRR----SIAGFVASINEGMTR 657
                ||             : || |    |:|||| |   |  |    |:|  ||:::
Db        716 GGINSRIVADKITNKYLVDQP---TMVVGIDVTHPTQAEMRMNMPSVAAIVANVDLLPQS 772

Qy        658 WFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEV 717
                : :   | : :| |  |   ::  :   :    |:|||||||| :||   ::  |:
Db        773 YGANVKVQKKCRESVVYLLDAIRERIITFYRHTKQKPARIIVYRDGVSEGQFSEVLREEI 832

Qy        718 PQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQ----SGGRLQNPLPGTVIDVEVTRPE 773
                    :|   : |  | |||:|| : | | :      |: :|  ||| :|  :  ||
Db        833 QSIRTACLAIAEDFRPPITYIVVQKRHHARIFCKYQNDMVGKAKNVPPGTTVDTGIVSPE 892

Qy        774 WYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQ 833
                 :||:: |     |:  |  |:|: |      | || :|| :|| |      : :| |
Db        893 GFDFYLCSHYGVQGTSRPARYHVLLDECKFTADEIQSITYGMCHTYGRCTRSVSIPTPVY 952

Qy        834 YAHKLAFLVGQSIHREPNLSLSN 856
                ||  :|     : |:  |: :|
```

Db 953 YADLVATRARCHVKRKLGLADNN 975

Search completed: July 22, 2003, 07:05:56
Job time : 20.1126 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2003, 22:53:12 ; Search time 10.0663 Seconds
(without alignments)
3547.609 Million cell updates/sec

Title: US-09-873-737A-6
Perfect score: 4534
Sequence: 1 MTGRARARARGRARGQETAQ.....VGQSIHREPNLSLSNRLYYL 861

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	539.5	11.9	834	1	IF2C_SCHPO	O74957 schizosacch
2	521	11.5	988	1	PINH_ARATH	Q9xgw1 arabidopsis
3	517	11.4	1048	1	AGO1_ARATH	O04379 arabidopsis
4	504.5	11.1	997	1	AGOL_ARATH	Q9sjk3 arabidopsis
5	496.5	11.0	863	1	I2C2_RAT	Q9qz81 rattus norv
6	495.5	10.9	857	1	I2C1_HUMAN	Q9ul18 homo sapien
7	474	10.5	813	1	I2C2_RABIT	O77503 oryctolagus
8	421	9.3	1040	1	YO43_CAEEL	P34681 caenorhabdi
9	371	8.2	377	1	I2C2_HUMAN	Q9ukv8 homo sapien
10	235.5	5.2	1032	1	YQ53_CAEEL	Q09249 caenorhabdi
11	117	2.6	1120	1	MFD_RICPR	O05955 rickettsia
12	116	2.6	1839	1	CYAA_SACKL	P23466 saccharomyc
13	115.5	2.5	2035	1	HMP2_YEREN	P48633 yersinia en
14	114	2.5	1653	1	CLH_YEAST	P22137 saccharomyc
15	111	2.4	1277	1	YMW6_YEAST	Q04264 saccharomyc
16	111	2.4	3144	1	VP13_YEAST	Q07878 saccharomyc
17	110.5	2.4	560	1	PHAC_PSEOL	P26496 pseudomonas
18	108	2.4	618	1	NXF1_MOUSE	Q99jx7 mus musculu
19	107.5	2.4	869	1	YD95_YEAST	P38966 saccharomyc
20	107	2.4	102	1	GGC1_HUMAN	O60829 homo sapien
21	106.5	2.3	2670	1	IP3T_RAT	Q63269 rattus norv
22	106	2.3	1317	1	N145_YEAST	P49687 saccharomyc
23	105.5	2.3	1597	1	RLR1_YEAST	P53552 saccharomyc
24	105	2.3	860	1	MUTS_PASMU	P57972 pasteurella
25	105	2.3	1180	1	C4AA_BACTI	P16480 bacillus th
26	105	2.3	1388	1	RPOD_TOBAC	P38550 nicotiana t
27	104	2.3	1384	1	RPOD_SINAL	Q9thv5 sinapis alb
28	103.5	2.3	1051	1	VP2_AHSV6	O71024 african hor
29	103.5	2.3	1170	1	ITA2_BOVIN	P53710 bos taurus
30	103	2.3	949	1	PMA6_ARATH	Q9sh76 arabidopsis
31	102.5	2.3	1066	1	HGPC_HAEIN	Q9x442 haemophilus
32	102	2.2	879	1	MYSP_DROME	P35415 drosophila
33	102	2.2	1202	1	DPOM_ASCIM	P22374 ascobolus i
34	102	2.2	1343	1	RPOB_HAEIN	P43738 haemophilus
35	101.5	2.2	878	1	MSH4_YEAST	P40965 saccharomyc
36	101.5	2.2	1017	1	HIRA_HUMAN	P54198 homo sapien
37	101.5	2.2	1208	1	PCP1_SCHPO	Q92351 schizosacch
38	101.5	2.2	1377	1	RPOC_BORBU	O51349 borrelia bu
39	101	2.2	1178	1	ITA2_MOUSE	Q62469 mus musculu
40	100.5	2.2	1357	1	RPOB_PSEAE	Q51561 pseudomonas
41	100	2.2	982	1	P115_MYCPN	P75361 mycoplasma
42	100	2.2	1575	1	IQG2_HUMAN	Q13576 homo sapien
43	99.5	2.2	648	1	GYRB_MYCHO	P43053 mycoplasma
44	99.5	2.2	1014	1	NEBL_HUMAN	O76041 homo sapien
45	99	2.2	934	1	KGP3_DROME	P32023 drosophila

ALIGNMENTS

RESULT 1
IF2C_SCHPO
ID IF2C_SCHPO STANDARD; PRT; 834 AA.
AC O74957;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable eukaryotic translation initiation factor 2C (eIF2C) (eIF-2C).
GN SPCC736.11.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC initiation process (By similarity).
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AL023705; CAA19275.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Initiation factor; Protein biosynthesis.
FT   DOMAIN      212    326       PAZ.
FT   DOMAIN      500    799       PIWI.
SQ   SEQUENCE   834 AA;  94438 MW;  F5CE397BF590D9E4 CRC64;

  Query Match           11.9%;  Score 539.5;  DB 1;  Length 834;
  Best Local Similarity  22.5%;  Pred. No. 4.4e-32;
  Matches  195;  Conservative  154;  Mismatches  350;  Indels  167;  Gaps   31;

Qy       106 GSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDL--------- 156
Db        16 GGLGKQITLKANFFQIISLPNETINQYHVIVGDGSRVPRKQSQLIWNSKEVKQYFGSSWM 75

Qy       157 ----------IGKCHAFDGTILF------LPKRLQ---QKVTEV-------FSKTRNGED 190
Db        76 NSVYDGRSMCWSKGDIADGTIKVNIGSESHPREIEFSIQKSSKINLHTLSQFVNSKYSSD 135

Qy       191 VRITITLTNELPPTSPTCLQFYNIIFRR-----LLKIMNLQQIGRNYYNPNDPIDIPSHR 245
Db       136 ------------PQVLSSIMFLDLLLKKKPSETLFGFMHSFFTGENGVSLGGGVE----- 178

Qy       246 LVIWPGFTTSILQYENSIMLCTDVSHKVL-RSETVLDFM---------------------- 283
Db       179 --AWKGFYQSIRPNQGFMSVNVDISSSAFWRNDSLLQILMEYTDCSNVRDLTRFDLKRLS 236

Qy       284 --FNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKK---ADGS 338
Db       237 RKFRFLKVTCQH-------RNNVG----TDLANRVYSIEGFSSKSASDSFFVRRLNGEEQ 285

Qy       339 GVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLT---GLTDK 395
Db       286 KISVAEYFLENHNVRLQYPNLPCIL------------VKNGAMLPIEFCFVVKGQRYTAK 333

Qy       396 MRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKND-NVQRELRDWGLSFDSNLLSFSG 454
Db       334 LNSD-----QTANMIRFAVQRPFERVQQIDDFVHQMDWDTDPYLTQYGMKIQKKMLEVPA 388

Qy       455 RILQTEKIHQGGGKTFDYNPQFADWSKETRGAPLISV--KPLDNWLLI---YTRR-NYEAA 508
Db       389 RVLETPSIRYGGDCIE-RPVSGRWN--LRGKRFLDPPRAPIRSWAVMCFTSTRRLPMRGI 445

Qy       509 NSLIQNLFKVTPAMGMQ--MRKAIMIEVDDR--TEAYLRVLQQKV----TADTQIVVCLL 560
Db       446 ENFLQTYVQTLTSLGINFVMKKPPVLYADIRGSVEELCITLYKKAEQVGNAPPDYLFFIL 505

Qy       561 SSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGG---ELW 617
Db       506 DKNSPEPYGSIKRVCNTMLGVPSQCAISKHI--LQSKPQYCANLGMKINVKVGGINCSLI 563

Qy       618 RVDIPLKLV--MIVGIDCYH-DMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDG 674
Db       564 PKSNPLGNVPTLILGGDVYHPGVGATGVSIASIVASVDLNGCKYTAVSRSQPRHQEVIEG 623

Qy       675 LKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPR 734
Db       624 MKDIVVYLLQGFRAMTKQQPQRIIYFRDGTSEGQFLSVINDELSQIKEACHSLSPKYNPK 683

Qy       735 XTVIVVKKRVNTRFFAQS---GGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSP 791
Db       684 ILVCTTQKRHHARFFIKNKSDGDRNGNPLPGTIIEKHVTHPYQYDFYLISHPSLQGVSVP 743

Qy       792 THYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFL---------- 841
Db       744 VHYTVLHDEIQMPPDQFQTLCYNLCYVYARATSAVSLVPPVYYAHLVSNLARYQDVTADD 803

Qy       842 ---------VGQSIHREPNLSLSNRL 858
Db       804 TFVETSEASMDQEV--KPLLALSSKL 827


RESULT 2
PINH_ARATH
ID   PINH_ARATH     STANDARD;      PRT;   988 AA.
AC   Q9XGW1; O49256;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   PINHEAD protein (ZWILLE protein).
GN   AT5G43810 OR MQD19.17.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Landsberg erecta;
RX   MEDLINE=99094873; PubMed=9876176;
RA   Lynn K., Fernandez A., Aida M., Sedbrook J., Tasaka M., Masson P.,
RA   Barton M.K.;
RT   "The PINHEAD/ZWILLE gene acts pleiotropically in Arabidopsis
RT   development and has overlapping functions with the ARGONAUTE1 gene.";
RL   Development 126:469-481(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Landsberg erecta;
RX   MEDLINE=98169396; PubMed=9501101;
RA   Moussian B., Schoof H., Haecker A., Juergens G., Laux T.;
RT   "Role of the ZWILLE gene in the regulation of central shoot meristem
RT   cell fate during Arabidopsis embryogenesis.";
RL   EMBO J. 17:1799-1809(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: DEVELOPMENTAL PROTEIN REQUIRED FOR RELIABLE FORMATION OF
CC       PRIMARY AND AXILLARY SHOOT APICAL MERISTEMS. MAY BE A COMPONENT OF
CC       A HYPOTHETICAL MERISTEM FORMING COMPETENCE FACTOR.
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AF154272; AAD40098.1; -.
DR   EMBL; AJ223508; CAA11429.1; -.
DR   EMBL; AB026651; BAB11310.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Developmental protein.
FT   DOMAIN      337    451       PAZ.
FT   DOMAIN      625    946       PIWI.
FT   CONFLICT    475    479       PRDRE -> AEGQR (IN REF. 2).
FT   CONFLICT    671    671       N -> D (IN REF. 2).
```

```
SQ   SEQUENCE   988 AA;  110867 MW;  32EBB349C613DA20 CRC64;

  Query Match             11.5%;  Score 521;  DB 1;  Length 988;
  Best Local Similarity   23.2%;  Pred. No. 1.3e-30;
  Matches  220;  Conservative  150;  Mismatches  424;  Indels  154;  Gaps   38;

Qy          4 RARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRGTAGGTAKSQ 63
              |:| | ||   |::: |        :|   :|| |||::                 |  |
Db         65 RSRRRNRG---GRKSDQ---GDVCMRPS-SRPRKPPPPSQ--------------TTSSA 102

Qy         64 GLQISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTS 123
                 :||  |:           | : :  |:| :       |: |    :  ||| |
Db        103 VSVATAG--EIVAVN--------HQMQMGVRKNSNFAPRPGFGTLGTKCIVKANHF-LAD 151

Qy        124 RPQWALYQYHIDYNPLMEARRLRSAL------LFQHEDLIGKCHAFD--------GTILF 169
               |   | ||  :   | : :: :  |:      |::  ||  :  |:|        | : |
Db        152 LPTKDLNQYDVTITPEVSSKSVNRAIIAELVRLYKESDLGRRLPAYDGRKSLYTAGELPF 211

Qy        170 LPKRLQQKVTE----VFSKTRNGEDVRITIT-------------LTNELPPTSPTCLQFY 212
               |    |: :    : :  :     :: |              |  :         :|
Db        212 TWKEFSVKIVDEDDGIINGPKRERSYKVAIKFVARANMHHLGEFLAGKRADCPQEAVQIL 271

Qy        213 NIIFRRLLKIMNLQQIGRNYYNPNDPIDIP---SHRLVIWPGFTTSILQYENSIMLCTDV 269
              :|: |  | :     :||::::|:  |  |       |  | ||  ||   :  : |  |:
Db        272 DIVLRE-LSVKRFCPVGRSFFSPD--IKTPQRLGEGLESWCGFYQSIRPTQMGLSLNIDM 328

Qy        270 -SHKVLRSETVLDFMFNFYHQ-------TEEHKFQEQVSKEXIGLVVLTKYNI-KTYRVD 320
               |   :    |::|:     :        ::  : : :     : : |  : |: : |||
Db        329 ASAAFIEPLPVIEFVAQLLGKDVLSKPLSDSDRVKIKKGLRGVKVEVTHRANVRRKYRVA 388

Qy        321 DIDWDQNPKSTFKKADGSGV-SFLEYYRKQYNQEITDLKQPVL-VSQPKRRRGPGGTLPG 378
              :      :  |   :   : | :||::: |    |     | | |   |:       ||
Db        389 GLTTQPTRELMFPVDENCTMKSVIEYFQEMYGFTIQHTHLPCLQVGNQKK----ASYLPM 444

Qy        379 PAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQREL 438
               |  | |    |   |  |:  :   | | |   |  |: :: | :   |   :
Db        445 EACKIVEGQRYT----KRLNEKQITALLKV-TCQRPRDRENDILRTVQ--HNAYDQDPYA 497

Qy        439 RDWGLSFDSNLLSFSGRILQTE--KIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNW 496
              :::|::    | |   |||     | |: || |  ||   |:   :   :|:   :  |
Db        498 KEFGMNISEKLASVEARILPAPWLKYHENGKEKDCLPQVGQWNMMNK--KMINGMTVSRW 555

Qy        497 LLIYTRRNYE--AANSLIQNLFKVTPAMGMQMRKAIMIEV----DDRTE-----AYLRVL 545
              :  |: :  |      | ::    ||:     :| :      |: |      |   :
Db        556 ACVNFSRSVQENVARGFCNELGQMCEVSGMEFNPEPVIPIYSARPDQVEKALKHVYHTSM 615

Qy        546 QQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVAR---TLGKQQTVMAIAT 602
              :    : :::: :|  |    |  :|:   |:    ||| : :    : ||
Db        616 NKTKGKELELLLAILPDNNGSLYGDLKRICETELGLISQCCLTKHVFKISKQ-----YLA 670

Qy        603 KIALQMNCKMGGE--------LWRV----DIPLKLVMIVGIDCYHDMTA--GRRSIAGFV 648
              ::|::| ||||           |:    |||    :| | |  |          |||  |
Db        671 NVSLKINVKMGGRNTVLVDAISCRIPLVSDIP---TIIFGADVTHPENGEESSPSIAAVV 727

Qy        649 ASIN-EGMTRWFSRCIFQDRGQELVDGLKVCLQAALR------------AWNSCNEYMP 694
              || :   :|::      |   |||:  |    |  :|             ::       |
Db        728 ASQDWPEVTKYAGLVCAQAHRQELIQDLYKTWQDPVRGTVSGGMIRDLLISFRKATGQKP 787

Qy        695 SRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFA---- 750
              ||| ||||| :||   :: ||:        |:   | |  | |||:|| :|| ||
Db        788 LRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEPNYQPPVTFIVVQKRHHTRLFANNHR 847

Qy        751 --QSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHI 808
                 |  |  | |||||:| ::  |  :||:: | |   |:  | ||:|:::| :    | |
Db        848 DKNSTDRSGNILPGTVVDTKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDENNFTADGI 907

Qy        809 QRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSN 856
              | ||  ||: |      : :  |  |||  ||      : || :   |
Db        908 QSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRA--RFYLEPEIMQDN 953


RESULT 3
AGO1_ARATH
```

```
ID   AGO1_ARATH     STANDARD;      PRT;  1048 AA.
AC   O04379;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Argonaute protein.
GN   AGO1 OR AT1G48410 OR F11A17.3.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia; TISSUE=Leaf;
RX   MEDLINE=98090460; PubMed=9427751;
RA   Bohmert K., Camus I., Bellini C., Bouchez D., Caboche M., Benning C.;
RT   "AGO1 defines a novel locus of Arabidopsis controlling leaf
RT   development.";
RL   EMBO J. 17:170-180(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).
CC   -!- FUNCTION: ESSENTIAL FOR PROPER DEVELOPMENT OF LEAVES AND FLORAL
CC       ORGANS, AND FORMATION OF AXILLARY MERISTEMS.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED AT LOW LEVELS.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT ALL DEVELOPMENTAL
CC       STAGES.
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U91995; AAC18440.1; -.
DR   EMBL; AC007932; AAD49755.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Developmental protein.
FT   DOMAIN      391    501       PAZ.
FT   DOMAIN      676    997       PIWI.
FT   DOMAIN       13    104       GLY-RICH.
```

```
SQ   SEQUENCE   1048 AA;  116190 MW;  3E5146343A09C541 CRC64;

  Query Match             11.4%;  Score 517;  DB 1;  Length 1048;
  Best Local Similarity   23.4%;  Pred. No. 2.9e-30;
  Matches  238;  Conservative  159;  Mismatches  428;  Indels  190;  Gaps   40;

Qy          1 MTGRARARARG---RARGQETAQLVGSTASQQP--------GYIQPRPQP---------- 39
              ::|  |   ||   :  ||       |   : ||        || ||  |
Db         25 VSGGGRGSQRGGFQQGGGQHQG---GRGYTPQPQQGGRGGRGYGQPPQQQQQYGGPQEYQ 81

Qy         40 ------PPAEGEL--FGRGRQRGTAGGTAKSQGLQ------------------------ 66
                  || :|    :| ||  | : |  : | :
Db         82 GRGRGGPPHQGGRGGYGGGRGGGPSSGPPQRQSVPELHQATSPTYQAVSSQPTLSEVSPT 141

Qy         67 -------ISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHF 119
                     ::  |::||  |:|   :    :  :::     ::  | | ||    :  |||
Db        142 QVPEPTVLAQQFEQLS-VEQGAPSQAIQPIPSSSKAFKFPMRPGK-GQSGKRCIVKANHF 199

Qy        120 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCH------AFDG-----TIL 168
                 |   |: | :   | : :| :  |:: | .|     |      |:||     |
Db        200 -FAELPDKDLHHYDVTITPEVTSRGVNRAVMKQLVDNYRDSHLGSRLPAYDGRKSLYTAG 258

Qy        169 FLP------------------KRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCL 209
               ||                  :| :::   |       :   : : |  :        |
Db        259 PLPFNSKEFRINLLDEEVGAGGQRREREFKVVIKLVARADLHHLGMFLEGKQSDAPQEAL 318

Qy        210 QFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHR-----LVIWPGFTTSILQYENSIM 264
               |  :|: |  |       :||:|:|    ||   :     |  | ||  ||   :  :
Db        319 QVLDIVLRE-LPTSRYIPVGRSFYSP----DIGKKQSLGDGLESWRGFYQSIRPTQMGLS 373

Qy        265 LCTDVSHKV-LRSETVLDFMFNFYHQ-------TEEHKFQEQVSKEXIGLVVLTKYNI-K 315
               |  |:|    :  :  |: |: :  ::       ::  : : : :   : : |  : |: :
Db        374 LNIDMSSTAFIEANPVIQFVCDLLNRDISSRPLSDADRVKIKKALRGVKVEVTHRGNMRR 433

Qy        316 TYRVDDIDWDQNPKSTFKKAD-GSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGG 374
               ||:  :       : ||   :  :  | :||: : |   |    : | |       |
Db        434 KYRISGLTAVATRELTFPVDERNTQKSVVEYFHETYGFRIQHTQLPCLQVGNSNR----- 488

Qy        375 TLPGPAMLIPELCYLT--GLTDKMRNDFNVMKDLAVHTRLTPEQRQREV---GRLIDYIH 429
                |  |  |:| :    .  |  |: :    | | |    |  |::::     :| ||
Db        489 ----PNYLPMEVCKIVEGQRYSKRLNERQITALLKV-TCQRPIDREKDILQTVQLNDY-- 541

Qy        430 KNDNVQRELRDWGLSFDSNLLSFSGRILQTE--KIHQGGKTFDYNPQFADWSKETRGAPL 487
               ||  :|   :|:   ::| |  |||      | |: |:     ||   |:   :   :
Db        542 AKDNYAQE---FGIKISTSLASVEARILPPPWLKYHESGREGTCLPQVGQWNMMNK--KM 596

Qy        488 ISVKPLDNWLLIYTRRNYE--AANSLIQNLFKVTPAMGMQMR-KAIMIEVDDRTEAYLRV 544
               |:   ::||: |   |  :   | :  | | ::    ||    : ::  |  | |   :|
Db        597 INGGTVNNWICINFSRQVQDNLARTFCQELAQMCYVSGMAFNPEPVLPPVSARPEQVEKV 656

Qy        545 LQQKVTADT---------QIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGK-Q 594
               |: :   |          ::: :|  |    |  :|:   |:    ||| : : : |
Db        657 LKTRYHDATSKLSQGKEIDLLIVILPDNNGSLYGDLKRICETELGIVSQCCLTKHVFKMS 716

Qy        595 QTVMAIATKIALQMNCKMGG-ELWRVD-----IPL---KLVMIVGIDCYHDMTA--GRRS 643
               :  ||    :||::| |:||     ||     |||   :  :| | |  |       |
Db        717 KQYMA---NVALKINVKVGGRNTVLVDALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPS 773

Qy        644 IAGFVASIN-EGMTRWFSRCIFQDRGQELVDGL------------KVCLQAALRAWNSC 689
               ||  ||| :   :|::      |   |||:  |                ::  | |:
Db        774 IAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFKEWKDPQKGVVTGGMIKELLIAFRRS 833

Qy        690 NEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFF 749
               : | ||| ||||| :||   :: ||:          |:  || |  | :||:|| :|| |
Db        834 TGHKPLRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEAGYQPPVTFVVVQKRHHTRLF 893

Qy        750 AQ------SGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGL 803
               ||      |  |  | ||||:| ::  |  :||:: | |   |:  | ||:|::| :
Db        894 AQNHNDRHSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDENNF 953

Qy        804 KPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRL 858
               | :| ||  ||: |      : :  |  |||  ||      : ||  | |  :
Db        954 TADGLQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRA--RFYMEPETSDSGSM 1006
```

```
RESULT 4
AGOL_ARATH
ID   AGOL_ARATH     STANDARD;      PRT;   997 AA.
AC   Q9SJK3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Argonaute-like protein At2g27880.
GN   AT2G27880 OR T1E2.20.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA   Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA   Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA   Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA   Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA   Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AC006929; AAD21514.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
FT   DOMAIN      359    471       PAZ.
FT   DOMAIN      638    958       PIWI.
SQ   SEQUENCE   997 AA;  111134 MW;  DA93A1446C422F31 CRC64;

  Query Match             11.1%;  Score 504.5;  DB 1;  Length 997;
  Best Local Similarity   22.7%;  Pred. No. 2.3e-29;
  Matches  219;  Conservative  153;  Mismatches  407;  Indels  186;  Gaps   36;

Qy          3 GRARARARGR---ARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRGTAGG- 58
              | : :  |||    ||: |  |   ||:| |                  || ||  |
Db         46 GGSVSAGRGRGNVGRGENTGDL---TATQVP------------VASAVSGGRGRGNIGDP 90

Qy         59 ----TAKSQGLQI-SAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKT-------- 105
                    : |: : : |:  :|     |       :       |:  :  :  : :
Db         91 TFSVASSSKTVSVASSSKEESKNTEVSETMSNLQITSTETKPEMTSLPPASSKAVTFPVR 150

Qy        106 ---GSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEA----RRLRSALLFQHED--L 156
                  |: |  | :  ||| |       || | :  || : :    | :  |:  ::|  |
Db        151 PGRGTLGKKVMVRANHF-LVQVADRDLYHYDVSINPEVISKTVNRNVMKLLVKNYKDSHL 209

Qy        157 IGKCHAFD--------GTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTC 208
                 ||  |:|        | : |  |      : |  :   :|:|    : : |    ||
Db        210 GGKSPAYDGRKSLYTAGPLPFDSKEFVVNLAEKRADGSSGKDRPFKVAVKN---VTSTDL 266
```

```
Qy        209 LQFYNIIFRR-------LLKIMNL----------QQIGRNYYNPNDPIDIPSHR------ 245
Db        267 YQLQQFLDRKQREAPYDTIQVLDVVLRDKPSNDYVSVGRSFFHTSLGKDARDGRGELGDG 326

Qy        246 LVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVL-DFMFNFYHQTEEHKFQEQVSKEXI 304
Db        327 IEYWRGYFQSLRLTQMGLSLNIDVSARSFYEPIVVTDFISKFLNIRDLNRPLRDSDR--- 383

Qy        305 GLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFL----------------EYY 346
Db        384 ---LKVKKVLRTLKVKLLHW-----NCTKSAKISGISSLPIRELRFTLEDKSEKTVVQYF 435

Qy        347 RKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELC-------YLTGLTDKMRND 399
Db        436 AEKYNYRVKYQALPAIQT---------GSDTRPVYLPMELCQIDEGQRYTKRLNEKQ--- 483

Qy        400 FNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQT 459
Db        484 ---VTALLKATCQRPPDRENSIKNLVVKNNYNDDLSKE---FGMSVTTQLASIEARVLPP 537

Qy        460 E--KIHQGGKTFDYNPQFADWS----KETRGAPLISVKPLDNWLLI--YTRRNYEAANSL 511
Db        538 PMLKYHDSGKEKMVNPRLGQWNMIDKKMVNGAKVTS------WTCVSFSTRIDRGLPQEF 591

Qy        512 IQNLFKVTPAMGMQMRKA-----IMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKD 566
Db        592 CKQLIGMCVSKGMEFKPQPAIPFISCPPEHIEEALLDI--HKRAPGLQLLIVIL-PDVTG 648

Qy        567 KYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGG------ELWRVD 620
Db        649 SYGKIKRICETELGIVSQCCQPRQVNKLN--KQYMENVALKINVKTGGRNTVLNDAIRRN 706

Qy        621 IPL---KLVMIVGIDCYHDMTA--GRRSIAGFVASIN-EGMTRWFSRCIFQDRGQELVDG 674
Db        707 IPLITDRPTIIMGADVTHPQPGEDSSPSIAAVVASMDWPEINKYRGLVSAQAHREEIIQD 766

Qy        675 LKVCLQAALR-------------AWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFL 721
Db        767 LYKLVQDPQRGLVHSGLIREHFIAFRRATGQIPQRIIFYRDGVSEGQFSQVLLHEMTAIR 826

Qy        722 DCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGG------RLQNPLPGTVIDVEVTRPEWY 775
Db        827 KACNSLQENYVPRVTFVIVQKRHHTRLFPEQHGNRDMTDKSGNIQPGTVVDTKICHPNEF 886

Qy        776 DFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYA 835
Db        887 DFYLNSHAGIQGTSRPAHYHVLLDENGFTADQLQMLTNNLCYTYARCTKSVSIVPPAYYA 946

Qy        836 HKLAF 840
Db        947 HLAAF 951
```

```
RESULT 5
I2C2_RAT
ID   I2C2_RAT       STANDARD;      PRT;   863 AA.
AC   Q9QZ81;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2)
DE   (Golgi ER protein 95 kDa) (GERp95).
GN   EIF2C2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Hepatoma;
RX   MEDLINE=99443791; PubMed=10512872;
RA   Cikaluk D.E., Tahbaz N., Hendricks L.C., DiMattia G.E., Hansen D.,
RA   Pilgrim D., Hobman T.C.;
RT   "GERp95, a membrane-associated protein that belongs to a family of
RT   proteins involved in stem cell differentiation.";
RL   Mol. Biol. Cell 10:3357-3372(1999).
CC   -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC       initiation process.
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF195534; AAF12800.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Initiation factor; Protein biosynthesis.
FT   DOMAIN      239    352       PAZ.
FT   DOMAIN      521    822       PIWI.
SQ   SEQUENCE   863 AA;  97659 MW;  9A1E0D6B84F7B622 CRC64;
```

Query Match 11.0%; Score 496.5; DB 1; Length 863;
Best Local Similarity 24.1%; Pred. No. 7.1e-29;
Matches 197; Conservative 149; Mismatches 351; Indels 121; Gaps 33;

```
Qy        106 GSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSAL---LFQH--EDLIG-K 159
Db         36 GTTGRTIKLQANFFEM-DIPKIDIYHYELDIKPEKCPRRVNREIVEHMVQHFKTQIFGDR 94

Qy        160 CHAFDG-----TILFLPKRLQQKVTEVFSKTRNGED--VRITI----------TLTN 199
Db         95 KPVFDGRKNLYTAMPLPIG-RDKVELEVTLPGEGKDRIFKVSIKWVSCVSLQALHDALSG 153

Qy        200 ELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIP-SHRLVIWPGFTTSILQ 258
Db        154 RLPSVPFETIQALDVVMRH-LPSMRYTPVGRSFFTASEGCSNPLGGGREVWFGFHQSVRP 212

Qy        259 YENSIMLCTDVSHKVL-RSETVLDF---MFNFYHQTEEHK-----FQEQVSKEXIGLVVL 309
Db        213 SLWKMMLNIDVSATAFYKAQPVIEFVCEVLDFKSIEEQQKPLTDSQRVKFTKEIKGLKVE 272

Qy        310 TKY---NIKTYRVDDIDWDQNPKSTFKKADGSG----VSFLEYYRKQYNQEITDLKQPVL 362
Db        273 ITHCGQMKRKYRVCNVTRRPASHQTFPLQQESGQTVECTVAQYFKDRHKLVLRYPHLPCL 332

Qy        363 -VSQPKRRRGPGGTLPGPAMLIPELCYLTG-------LTDKMRNDFNVMKDLAVHTRLTP 414
Db        333 QVGQEQKH----------TYLPLEVCNIVAGQRCIKKLTD---NQTSTM--IRATARSAP 377

Qy        415 EQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQ 474
Db        378 D-RQEEISKLMR--SASFNTDPYVREFGIMVKDEMTDVTGRVLQPPSILYGGR------- 427

Qy        475 FADWSKETRGAPLISVKPLDN----------WLLIYTRRNYEAA----NSLIQNLFKVT 519
Db        428 -----NKAIATPVQGVWDMRNKQFHTGIEIKVWAIACFAPQRQCTEVHLKSFTEQLRKIS 482

Qy        520 PAMGMQMRK----AIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYL 575
Db        483 RDAGMPIQGQPCFCKYAQGADSVEPMFRHL-KNTYAGLQLVVVIL-PGKTPVYAEVKRVG 540

Qy        576 CTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGG------ELWRVDIPLKLVMIV 629
Db        541 DTVLGMATQCVQMKNV--QRTTPQTLSNLCLKINVKLGGVNNILLPQGRPPVFQQPVIFL 598
```

```
Qy        630 GIDCYHDMTAG---RRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAW 686
Db        599 GADVTHP-PAGDGKKPSIAAVVGSMDAHPNRYCATVRVQQHRQEIIQDLAAMVRELLIQF 657

Qy        687 NSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNT 746
Db        658 YKSTRFKPTRIIFYRDGVSEGQFQQVLHHELLAIREACIKLEKEYQPGITFIVVQKRHHT 717

Qy        747 RFFA----QSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSG 802
Db        718 RLFCTDKNERVGKSGNIPAGTTVDTKITHPTEFDFYLCSHAGIQGTSRPSHYHVLWDDNR 777

Qy        803 LKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAF 840
Db        778 FSSDELQILTYQLCHTYVRCTRSVSIPAPAYYAHLVAF 815
```

RESULT 6
I2C1_HUMAN
ID I2C1_HUMAN STANDARD; PRT; 857 AA.
AC Q9UL18;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Eukaryotic translation initiation factor 2C 1 (eIF2C 1) (eIF-2C 1)
DE (Putative RNA-binding protein Q99).
GN EIF2C1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20005943; PubMed=10534406;
RA Koesters R., Adams V., Betts D., Moos R., Schmid M., Siermann A.,
RA Hassam S., Weitz S., Lichter P., Heitz P.U., von Knebel Doeberitz M.,
RA Briner J.;
RT "Human eukaryotic initiation factor EIF2C1 gene: cDNA sequence,
RT genomic organization, localization to chromosomal bands 1p34-p35, and
RT expression.";
RL Genomics 61:210-218(1999).
CC -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC initiation process.
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; AF093097; AAF00068.1; -.
DR Genew; HGNC:3262; EIF2C1.
DR MIM; 606228; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
KW Initiation factor; Protein biosynthesis.
FT DOMAIN 226 346 PAZ.
FT DOMAIN 515 816 PIWI.
SQ SEQUENCE 857 AA; 97214 MW; 1DBB524AE7CBAF66 CRC64;

Query Match 10.9%; Score 495.5; DB 1; Length 857;
Best Local Similarity 24.2%; Pred. No. 8.4e-29;
Matches 204; Conservative 157; Mismatches 349; Indels 133; Gaps 37;

```
Qy        106 GSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSAL---LFQH--EDLIG-K 159
Db         30 GTVGKPIKLLANYFEV-DIPKIDVYHYEVDIKPDKCPRRVNREVVEYMVQHFKPQIFGDR 88

Qy        160 CHAFDG-----TILFLP---KRLQQKVTEVFSKTRNGED--VRITI----------T 196
Db         89 KPVYDGKKNIYTVTALPIGNERVDFEVT----IPGEGKDRIFKVSIKWLAIVSWRMLHEA 144

Qy        197 LTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIP-SHRLVIWPGFTTS 255
Db        145 LVSGQIPVPLESVQALDVAMRHLAS-MRYTPVGRSFFSPPEGYYHPLGGGREVWFGFHQS 203

Qy        256 ILQYENSIMLCTDVSHKVL-RSETVLDFMF------NFYHQTEEHKFQEQV--SKEXIGL 306
Db        204 VRPAMWKMMLNIDVSATAFYKAQPVIEFMCEVLDIRNIDEQPKPLTDSQRVRFTKEIKGL 263

Qy        307 VVLTKY---NIKTYRVDDIDWDQNPKSTFKKADGSG----VSFLEYYRKQYNQEITDLKQ 359
Db        264 KVEVTHCGQMKRKYRVCNVTRRPASHQTFPLQLESGQTVECTVAQYFKQKYNLQ---LKY 320

Qy        360 PVL----VSQPKRRRGPGGTLPGPAMLIPELCYLTG-------LTDKMRNDFNVMKDLAV 408
Db        321 PHLPCLQVGQEQKH----------TYLPLEVCNIVAGQRCIKKLTD---NQTSTM--IKA 365

Qy        409 HTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKT 468
Db        366 TARSAPD-RQEEISRLMK--NASYNLDPYIQEFGIKVKDDMTEVTGRVLPAPILQYGGRN 422

Qy        469 FDY-NPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQN----LFKVTPAMG 523
Db        423 RAIATPNQGVW--DMRGKQFYNGIEIKVWAIACFAPQKQCREEVLKNFTDQLRKISKDAG 480

Qy        524 MQMRK----AIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDC 579
Db        481 MPIQGQPCFCKYAQGADSVEPMFRHLKNTYSG-LQLIIVIL-PGKTPVYAEVKRVGDTLL 538

Qy        580 PTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIP-------LKLVMIVGID 632
Db        539 GMATQCVQVKNVVK--TSPQTLSNLCLKINVKLGG-INNILVPHQRSAVFQQPVIFLGAD 595

Qy        633 CYHDMTAG---RRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSC 689
Db        596 VTHP-PAGDGKKPSITAVVGSMDAHPSRYCATVRVQRPRQEIIEDLSYMVRELLIQFYKS 654

Qy        690 NEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFF 749
Db        655 TRFKPTRIIFYRDGVPEGQLPQILHYELLAIRDACIKLEKDYQPGITYIVVQKRHHTRLF 714

Qy        750 A----QSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKP 805
Db        715 CADKNERIGKSGNIPAGTTVDTNITHPFEFDFYLCSHAGIQGTSRPSHYYVLWDDNRFTA 774

Qy        806 DHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAF-----------------LVGQSIH 847
Db        775 DELQILTYQLCHTYVRCTRSVSIPAPAYYARLVAFRARYHLVDKEHDSGEGSHISGQSNG 834

Qy        848 REP 850
Db        835 RDP 837
```

RESULT 7
I2C2_RABIT
ID I2C2_RABIT STANDARD; PRT; 813 AA.
AC O77503;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2).
GN EIF2C2.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Liver;
RX   MEDLINE=98267198; PubMed=9602122;
RA   Zou C., Zhang Z., Wu S., Osterman J.C.;
RT   "Molecular cloning and characterization of a rabbit eIF2C protein.";
RL   Gene 211:187-194(1998).
CC   -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC       initiation process.
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF005355; AAC24323.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Initiation factor; Protein biosynthesis.
FT   DOMAIN      189    302       PAZ.
FT   DOMAIN      471    772       PIWI.
SQ   SEQUENCE   813 AA;  92281 MW;  13BA0DB70E10688F CRC64;

  Query Match             10.5%;  Score 474;  DB 1;  Length 813;
  Best Local Similarity   23.9%;  Pred. No. 3.1e-27;
  Matches  191;  Conservative  145;  Mismatches  343;  Indels  120;  Gaps   32;

Qy        125 PQWALYQYHIDYNPLMEARRLRSAL---LFQH--EDLIG-KCHAFDG-----TILFLPKR 173
              |:  :| | :|  |    ||:   :   : ||    : | :   |||     | : ||
Db          4 PKIDIYHYELDIKPEKCPRRVNREIVEHMVQHFKAQIFGDRKPVFDGRKNLYTAMPLPIG 63

Qy        174 LQQKVTEVFSKTRNGED--VRITI------------TLTNELPPTSPTCLQFYNIIFRR 218
               ::||    :    |:|   :::|              |:  ||      :|  ::: |
Db         64 -REKVELEVTLPGEGKDRIFKVSIKWVSCVSLQALHDALSGRLPSVPFETIQALDVVMRH 122

Qy        219 LLKIMNLQQIGRNYYNPNDPIDIP-SHRLVIWPGFTTSILQYENSIMLCTDVSHKVL-RS 276
               |  |    :||:::  ::   |    :| ||  |:      :||  |||     ::
Db        123 -LPSMRYTPVGRSFFTASEGCSNPLGGGREVWFGFHQSVRPSLWKMMLNIDVSATAFYKA 181

Qy        277 ETVLDF---MFNFYHQTEEHK-----FQEQVSKEXIGLVVLTKY---NIKTYRVDDIDWD 325
              : |::|   : :|    |: |      : : :||  || |   :     : ||| ::
Db        182 QPVIEFVCEVLDFKSIEEQQKPLTDSQRVKFTKEIKGLKVEITHCGQMKRKYRVCNVTRR 241

Qy        326 QNPKSTFKKADGSG----VSFLEYYRKQYNQEITDLKQPVL-VSQPKRRRGPGGTLPGPA 380
                  ||     ||     :  :|:: ::   :     | | | | ::
Db        242 PASHQTFPLQQESGQTVECTVAQYFKDRHKLVLRYPHLPCLQVGQEQKH----------T 291

Qy        381 MLIPELCYLTG-------LTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDN 433
               |  |:| :         |||   |  : |  :     |  |: || |: :|:     : |
Db        292 YLPLEVCNIVAGQRCIKKLTD---NQTSTM--IRATARSAPD-RQEEISKLMR--SASFN 343

Qy        434 VQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPL 493
                  :|::|:     :   :||:||   |  ||:              :    |:  |  :
Db        344 TDPYVREFGIMVKDEMTDVTGRVLQPPSILYGGR-----------NKAIATPVQGVWDM 391

Qy        494 DN----------WLLIYTRRNYEAA----NSLIQNLFKVTPAMGMQMRK----AIMIEV 534
               |          | :       :       |  : | |::   || ::         :
Db        392 RNKQFHTGIEIKVWAIACFAPQRQCTEVHLKSFTEQLRKISRDAGMPIQGQPCFCKYAQG 451

Qy        535 DDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQ 594
               |      | | :   |  |:|| :|   :   |  :|:   |     :|||  : :  |
```

```
Db        452 ADSVGPMFRHL-KNTYAGLQLVVVIL-PGKTPVYAEVKRVGDTVLGMATQCVQMKNV--Q 507

Qy        595 QTVMAIATKIALQMNCKMGG------ELWRVDIPLKLVMIVGIDCYHDMTAG---RRSIA 645
              :|     : : |::| |:||          |  :  : |: :| |  |   ||   : |||
Db        508 RTTPQTLSNLCLKINVKLGGVNNILLPQGRPPVFQQPVIFLGADVTHP-PAGDGKKPSIA 566

Qy        646 GFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVG 705
                | |::    |: :    |   ||::  |   ::  |  :     : |:||| |||||
Db        567 AVVGSMDAHPNRYCATVRVQQHRQEIIQDLAAMVRELLIQFYKSTRFKPTRIIFYRDGVS 626

Qy        706 DGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFA----QSGGRLQNPLP 761
              :|| : ::::|:    :    : : | |  | |||:|| :|| |     :  |:  |
Db        627 EGQFQQVLHHELLAIREACIKLEKDYQPGITFIVVQKRHHTRLFCTDKNERVGKSGNIPA 686

Qy        762 GTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYN 821
              || :| ::| |  :||:: | |   |:  |:||:|::|::    | :| |||:||| |
Db        687 GTTVDTKITHPTEFDFYLCSHAGIQGTSRPSHYHVLWDDNRFSSDELQILTYQLCHTYVR 746

Qy        822 WPGVIRVPAPCQYAHKLAF 840
                : :|||  ||| :||
Db        747 CTRSVSIPAPAYYAHLVAF 765


RESULT 8
YO43_CAEEL
ID   YO43_CAEEL     STANDARD;      PRT;  1040 AA.
AC   P34681;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical 115.4 kDa protein ZK757.3 in chromosome III.
GN   ZK757.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA   Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA   Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA   Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA   Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA   Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA   Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z30215; CAA82941.1; -.
DR   EMBL; Z29121; CAA82941.1; JOINED.
DR   EMBL; Z29121; CAA82389.1; -.
DR   EMBL; Z30215; CAA82389.1; JOINED.
DR   PIR; S41013; S41013.
DR   WormPep; ZK757.3; CE01117.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
```

```
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Hypothetical protein.
FT   DOMAIN      378    486       PAZ.
FT   DOMAIN      660    966       PIWI.
SQ   SEQUENCE   1040 AA;  115415 MW;  4C8483C1F1D72338 CRC64;

  Query Match             9.3%;  Score 421;  DB 1;  Length 1040;
  Best Local Similarity   21.6%;  Pred. No. 3.8e-23;
  Matches  199;  Conservative  148;  Mismatches  364;  Indels  212;  Gaps   37;

Qy        106 GSSGIIVRLSTNHFRL-TSRPQWALYQYHID-YNPLMEARRL---RSALLF-----QHED 155
              |: |  : : :| | :    |:  : ||||:: ::|    |:|      ::|      | :
Db         93 GTIGRQIPVKSNFFAMDLKNPKMVVIQYHVEIHHP--GCRKLDKDEMRIIFWKAVSDHPN 150

Qy        156 LIGK--CHAFDGT-ILFLPKRLQ--------QKVTEVFSKTRNGEDVRITI--------- 195
              :       |:||   |:   ||:        :   |       | :  |  |
Db        151 IFHNKFALAYDGAHQLYTVARLEFPDDQGSVRLDCEASLPKDNRDRTRCAISIQNVGPVL 210

Qy        196 -----TLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYN----------PNDPI 239
                  | || |     | :|  :|| |: |    |:    |:|                :
Db        211 LEMQRTRTNNLDERVLTPIQILDIICRQSLTCPLLKN-SANFYTWKSSCYRIPTAAGQAL 269

Qy        240 DIPSHRLVIWPGFTTSILQYEN-SIMLCTDVSHKVLRSE--TVLDFMFNFYHQ------- 289
              |:   :  :| || :|     |   :|  ||:|        ||| || :  ::
Db        270 DLEGGK-EMWTGFFSSAHIASNYRPLLNIDVAHTAFYKTRITVLQFMCDVLNERTSKPNR 328

Qy        290 ----------------------------------------------------------T 290
                                                                           |
Db        329 NNPRGPGGPGGPGGYRGGRGGGRGGSYGNFGNRGPPGANVRDDFGGNGLTFTMDTLSRDT 388

Qy        291 EEHKFQEQVSKEXI-GLVVLTKYN---IKTYRVDDIDWDQNP--KSTFKKADGSG----V 340
              :   |: ::  : | |: :   :    |: |:|: :   | |  |  |:  |  |
Db        389 QLSSFETRIFGDAIRGMKIRAAHRPNAIRVYKVNSL---QLPADKLMFQGIDEEGRQVVC 445

Qy        341 SFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIP-ELC-------YLTGL 392
              |  :|: ::|      || | |   |    ||    |   : :| | |       |   :
Db        446 SVADYFSEKYG----PLKYPKL---PCLHVGP----PTRNIFLPMEHCLIDSPQKYNKKM 494

Qy        393 TDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSF 452
              ::|  :   ::|  ||       ||:  : :|      :      |:::|::  | ::
Db        495 SEKQTS--AIIKAAAVDA----TQREDRIKQLA--AQASFGTDPFLKEFGVAVSSQMIQT 546

Qy        453 SGRILQTEKIHQGGKTFDYN----PQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA 508
              : |::|   |  ||      |    |:   |: : :   : :      : : :   |:  :
Db        547 TARVIQPPPIMFGGNNRSVNPVVFPKDGSWTMDNQTLYMPATCRSYSMIALVDPRDQTSL 606

Qy        509 NSLIQNLFKVTPAMGMQMRK--------------AIMIEVDDRTEAYLRVLQQKVTADT 553
              :  |:|     ||||   :              :  |: |          : :||
Db        607 QTFCQSLTMKATAMGMNFPRWPDLVKYGRSKEDVCTLFTEIAD---------EYRVTNTV 657

Qy        554 -QIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKM 612
               :: :| |   | |  :|:          ||||: : : :      |    | |::| ||
Db        658 CDCIIVVLQSKNSDIYMTVKEQSDIVHGIMSQCVLMKNVSR--PTPATCANIVLKLNMKM 715

Qy        613 GG-----------ELWRVDIPLKLVMIVGIDCYHDMTAGRR----SIAGFVASINEGMTR 657
              ||             : || |    |:|||| |   |  |    |:|  ||:::
Db        716 GGINSRIVADKITNKYLVDQP---TMVVGIDVTHPTQAEMRMNMPSVAAIVANVDLLPQS 772

Qy        658 WFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEV 717
              : :    | : :| | |   ::  :  :    :  |:||||||||| :||   ::  |:
Db        773 YGANVKVQKKCRESVVYLLDAIRERIITFYRHTKQKPARIIVYRDGVSEGQFSEVLREEI 832

Qy        718 PQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQ----SGGRLQNPLPGTVIDVEVTRPE 773
                    :|   : |  | |||:|| : | | :      :|: :|  ||| :|  :  ||
Db        833 QSIRTACLAIAEDFRPPITYIVVQKRHHARIFCKYQNDMVGKAKNVPPGTTVDTGIVSPE 892

Qy        774 WYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQ 833
              :||:: |     |:  |  |:|: |      | || :|| :|| |      : :| |
Db        893 GFDFYLCSHYGVQGTSRPARYHVLLDECKFTADEIQSITYGMCHTYGRCTRSVSIPTPVY 952

Qy        834 YAHKLAFLVGQSIHREPNLSLSN 856
              ||  :|      : |:  |: :|
Db        953 YADLVATRARCHVKRKLGLADNN 975


RESULT 9
I2C2_HUMAN
ID   I2C2_HUMAN     STANDARD;      PRT;   377 AA.
AC   Q9UKV8;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2)
DE   (Fragment).
GN   EIF2C2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20005943; PubMed=10534406;
RA   Koesters R., Adams V., Betts D., Moos R., Schmid M., Siermann A.,
RA   Hassam S., Weitz S., Lichter P., Heitz P.U., von Knebel Doeberitz M.,
RA   Briner J.;
RT   "Human eukaryotic initiation factor EIF2C1 gene: cDNA sequence,
RT   genomic organization, localization to chromosomal bands 1p34-p35, and
RT   expression.";
RL   Genomics 61:210-218(1999).
CC   -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC       initiation process.
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF121255; AAF13034.2; -.
DR   Genew; HGNC:3263; EIF2C2.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Initiation factor; Protein biosynthesis.
FT   NON_TER       1      1
FT   DOMAIN       35    336       PIWI.
SQ   SEQUENCE   377 AA;  42529 MW;  3BBCE91689D58526 CRC64;

  Query Match             8.2%;  Score 371;  DB 1;  Length 377;
  Best Local Similarity   30.2%;  Pred. No. 4.3e-20;
  Matches   96;  Conservative   63;  Mismatches  141;  Indels   18;  Gaps    7;

Qy        536 DRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQ 595
              |  |   | | :   |  |:|| :|   :   |  :|:   |     :||| : :  |:
Db         17 DSVEPMFRHL-KNTYAGLQLVVVIL-PGKTPVYAEVKRVGDTVLGMATQCVQMKNV--QR 72

Qy        596 TVMAIATKIALQMNCKMGG------ELWRVDIPLKLVMIVGIDCYHDMTAG---RRSIAG 646
              |     : : |::| |:||         |  :  : |: :| |  |   ||   : |||
Db         73 TTPQTLSNLWLKINVKLGGVNNILLPQGRPPVFQEPVIFLGADVTHP-PAGDGKKPSIAA 131

Qy        647 FVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGD 706
              |  ::    |: :    |   ||::  |   ::  |  :      : |:||| |||||  :
Db        132 VVGRMDAHPNRYCATVRVQQHRQEIIQDLAAMVRELLIQFYKSTRFKPTRIIFYRDGVSE 191

Qy        707 GQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFA----QSGGRLQNPLPG 762
              || : ::::|:    :    : : | |  | |||:|| :|| |     :  |:  |   |
Db        192 GQFQQVLHHELLAIREACIKLEKDYQPGITFIVVQKRHHTRLFCTDKNERVGKSGNIPAG 251

Qy        763 TVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNW 822
```

```
                 | :| ::| |  :||:: | |   |:  |:||:|::|::    | :| |||:||| |
Db        252 TTVDTKITHPTEFDFYLCSHAGIQGTSRPSHYHVLWDDNRFSSDELQILTYQLCHTYVRC 311

Qy        823 PGVIRVPAPCQYAHKLAF 840
                : :|||  ||| :||
Db        312 TRSVSIPAPAYYAHLVAF 329


RESULT 10
YQ53_CAEEL
ID   YQ53_CAEEL     STANDARD;      PRT;  1032 AA.
AC   Q09249;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein C16C10.3 in chromosome III.
GN   C16C10.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Lloyd C.;
RL   Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: STRONG, TO C.ELEGANS C14B1.7. ALSO TO C06A1.4,
CC       F48F7.1, R06C7.1, T22B3.2 AND ZK757.3.
CC   -!- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PIWI DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z46787; CAA86741.1; -.
DR   WormPep; C16C10.3; CE01494.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
KW   Hypothetical protein.
FT   DOMAIN      369    481       PAZ.
FT   DOMAIN      650    977       PIWI.
SQ   SEQUENCE   1032 AA;  117105 MW;  EBEAE23CE67BB149 CRC64;

  Query Match             5.2%;  Score 235.5;  DB 1;  Length 1032;
  Best Local Similarity   20.6%;  Pred. No. 2.2e-09;
  Matches  129;  Conservative   98;  Mismatches  231;  Indels  169;  Gaps   24;

Qy        292 EHKFQEQ------VSKEXIGL-VVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSG---VS 341
              || | |:      : :   || |  |       ||  | :    ::|   |  |   |:
Db        381 EHAFGERGNTKRCIEEALKGLDVECTHLKGNLIRVSSIAENNAENTSFMMKDDKGEREVT 440

Qy        342 FLEYYRKQYNQEITDLKQPVLVSQPKR----------RRGPGGTLPGPAMLIPELCYLTG 391
               ||:  ||| ::   : |::||:  :          |  ||  :    |       :||
Db        441 VAEYFLLQYNIKLKYPRLPLVVSKRFKHESFFPMELLRIAPGQRIKVNKMSPTVQSAMTG 500

Qy        392 LTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLS 451
                |                 |:   : |  ::      ||:: |   :  :|   |:|
Db        501 RNASM----------------PQHHVKLVQDIL-----RDNLKLEQNKYMDAFGIKLMS 538

Qy        452 -----FSGRILQTEKIHQGGKTF--DYN-PQFADWSKETRGAPL--ISVKPLDNWLLI-- 499
                  : ::|   :|   |:|:  | : | |     |    | :  | :   || : :
Db        539 TEPIQMTAKLLPPAQIKFKGQTYMPDMSRPAFRTQDKFVEPARIRKIGIVVFDNCIQMRQ 598

Qy        500 ------------------------YTRRNYEAANSL-IQNLFKVTPAMGMQMRKAIMIE 533
                                      :: |   :::|: |||| |
Db        599 AEDFCDKLSNFCRDNGITVEKDSRDWSIRELNSSDSVAIQNLMKKW------------- 644

Qy        534 VDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKY----------LCTDCPTPS 583
              :||| :  : : ::|               : | :| :| :          ||
Db        645 LDDRVDILVGIAREK--------------KPDVHDILKYFEESIGLQTIQLC------- 682

Qy        584 QCVVARTLGKQ---QTVMAIATKIALQMNCKMGGELWRVDIP----------------- 622
              |  | : :| |   ||:  :  |     | | ||  : |:||
Db        683 QQTVDKMMGGQGGRQTIDNVMRK----FNLKCGGTNFFVEIPNAVRGKAVCSNNETLRKK 738

Qy        623 -LKLVMIVGIDCYHDMTA----------GRRSIAGFVASINEGMTRWFSRCIFQDRGQE 670
               |: |  :| :  |  :            |  |: |   |:    |:       | : :
Db        739 LLEHVQFIGFEISHGASRTLFDRSRSQMDGEPSVVGVSYSLTNS-TQLGGFTYLQTQKEY 797

Qy        671 LVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRG 730
               :  |       :|::   :: :|:||::|| | |:|     |  || :       | |
Db        798 KLQKLDEFFPKCVRSYKEHSKTLPTRIVIYRVGAGEGNFNR-VKEEVEEMRRTFDKIQPG 856

Qy        731 YNPRXTVIVVKKRVNTRFFAQ--SGGRL--QNPLPGTVIDVEVTRPEWYDFFIVSQAVRS 786
              | |   ||: ::  : | |    || |   ||   || ::  :|   : :| : ||
Db        857 YRPHLVVIIAQRASHARVFPSCISGNRATDQNIPSGTCVENVLTSYGYDEFILSSQTPLI 916

Qy        787 GSVSPTHYNVIYDNSGLKPDHIQRLTY 813
              |:| |  | :: :::    : :  |||
Db        917 GTVRPCKYTILVNDAKWSKNELMHLTY 943


RESULT 11
MFD_RICPR
ID   MFD_RICPR      STANDARD;      PRT;  1120 AA.
AC   O05955;
DT   15-DEC-1998 (Rel. 37, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Transcription-repair coupling factor (TRCF).
GN   MFD OR RP598.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=782;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Madrid E;
RX   MEDLINE=99039499; PubMed=9823893;
RA   Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA   Sicheritz-Ponten T., Alsmark U.C.M., Podowski·R.M., Naeslund A.K.,
RA   Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
RN   [2]
RP   SEQUENCE OF 132-1120 FROM N.A.
RC   STRAIN=Madrid E;
RX   MEDLINE=97419517; PubMed=9274032;
RA   Andersson J.O., Andersson S.G.E.;
RT   "Genomic rearrangements during evolution of the obligate
RT   intracellular parasite Rickettsia prowazekii as inferred from an
RT   analysis of 52015 bp nucleotide sequence.";
RL   Microbiology 143:2783-2795(1997).
CC   -!- FUNCTION: NECESSARY FOR STRAND-SPECIFIC REPAIR. A LESION IN THE
CC       TEMPLATE STRAND BLOCKS THE RNA POLYMERASE COMPLEX (RNAP). THE
CC       RNAP-DNA-RNA COMPLEX IS SPECIFICALLY RECOGNIZED BY TRCF WHICH
CC       RELEASES RNAP AND THE TRUNCATED TRANSCRIPT; THE TCRF MAY REPLACE
CC       RNAP AT THE LESION SITE AND THEN RECRUIT THE UVRA/B/C REPAIR
CC       SYSTEM (BY SIMILARITY).
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE UVRB FAMILY.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE HELICASE
CC       FAMILY. RECG SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AJ235272; CAA15042.1; -.
DR   EMBL; Y11783; CAA72474.1; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004576; Mfd.
DR   InterPro; IPR005118; TRCF.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF03461; TRCF; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   TIGRFAMs; TIGR00580; mfd; 1.
KW   Helicase; DNA repair; ATP-binding; DNA-binding; Complete proteome.
FT   NP_BIND     604    611       ATP (POTENTIAL).
FT   SITE        709    712       DEEQ BOX.
SQ   SEQUENCE   1120 AA;  127675 MW;  09C78EBE71D00482 CRC64;

  Query Match             2.6%;  Score 117;  DB 1;  Length 1120;
  Best Local Similarity   17.8%;  Pred. No. 1.5;
  Matches  144;  Conservative  131;  Mismatches  298;  Indels  234;  Gaps   36;

Qy       141 EARRL-RSALLFQHEDLIGKCHAFDGTILF--------LPKRLQQKVTEVFSKTRNGEDV 191
                 || :| : || |   | |    ::| || :         :  |  : :|:: : ::    :
Db        36 EALQLYKQALFFSSNDNIYYFPSYD-TIPYDHTSPNANIVSRRAETLTKLITNSKGKLLI 94

Qy       192 RITITLTNELPPTSPTCLQFYNIIFRRL-------LKIMNLQQIGRNYYNPNDPIDIPSH 244
                 | |:|||       |::  | :|       :  :::  :  ::       ||:
Db        95 THAANLLNKLPPK-----DFFSKYFLKLYPKIKFTMDELSMLLVENSFTRNTSSIDVGEF 149

Qy       245 RL------VIWPG-----------FTTSILQYENSIMLCT---------DVSHKVLRSET 278
               :     :| ||           :  || ::: :  : |           :  || |||
Db       150 AVRGEIIDIILPGPKGYRINFSWDYIESIKEFDINTQISTKYCAELVISPANEIVLNSET 209

Qy       279 VLDFMFNF-----YHQTEEHKFQEQVS-KEXIG---LVVLTKYNIKTYRVDDIDWDQNPK 329
               : :|  |:      : |:   ::  :| ::   |   |: |  |:  :     :|:  ||
Db       210 IGNFKNNYLRNFGVNHTDNPLYEAVISGRKFAGYEQLLPLFYYSCSSL----VDYLNNPI 265

Qy       330 STFKKADGSGV-----SFLEYY--RKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAML 382
               |       :     |: ::|  | : |:   :  |                      |
Db       266 CIFDNLSKQAILEFENSYNDFYLARSKANKLKVNNFYPT-------------------L 305

Qy       383 IPELCYLT--GLTDKMRNDFNVMKDLAVHTRLTPEQRQRE--VGRL--IDYIHKNDNVQR 436
               |   | |   :|: :    |::        :: |  ::   :| :  | :| |     :
Db       306 SPTSLYFTASAITELLEQKNNIL--------ISYENSEQASLIGNISSISFIEKKTIFDK 357

Qy       437 ELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPL-ISVKPLD- 494
                    :|   ::  | :   |:|     : ::|      :   | : : : | : ||:
Db       358 LFEIIKANFHKKIIICSSVLSSFERIKSIIQNYEYTFNEINKLDEAKASVINIGIIPLNQ 417

Qy       495 -----NWLLI-----------YTRRNYEAANSLIQ--NL----FKVTPAMGM-QMRKAIM 531
                   :| |            ||  | :  | |::  |:    | |    |: |  |
Db       418 SFYTKEYLFITSSELLEEKTLYTNTNKKLKNILLELDNLKEGEFVVHKDHGIGQFLKLEA 477

Qy       532 IEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRK-----DKYDAIKKYLCTDCPTPSQCV 586
               ::  :  :|::                |  | |     :  : ||||   :
Db       478 FKIQGKLHDFLKI--------------LYFGNDKLYVPVENIEVIKKYGSDNA------- 516

Qy       587 VARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMI-------VGIDCYHDMTA 639
                 | |  :|    :|  |:   |         :| | |: |       :       |: |
Db       517 ---ELNKLGSVAWNKSKAKLKNRIK--------EISLHLIQIAAKRKLNISTPIELDLEA 565

Qy       640 GRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGL--------------KVCLQAALR 684
               :  | | |  |  |             |  ::| :|:                :| ::|
Db       566 YDKFCANFPFSETEDQL-----TAINDIREDLTNGMLMDRLICGDVGFGKTEVAMRAVFM 620

Qy       685 AWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKK-- 742
               | |||:|   :|    :       | :    :|::  |  |       :||  |:
Db       621 VAKSLNEYLPQVAVVVPTTI-------LCSQHFSRFIERFKGFGLNIKQLSSVISSKEAN 673

Qy       743 --------------------RVNTRFFAQSGGRLQNPLPGTVIDVE----VTRPEWYDF 777
                               ||:||          |   :|| |    |:: |:
Db       674 IIRLELASGKINIIIGTHALLHKNTKFF---------NLKLLIIDEEQHFGVSQKEFLKS 724

Qy       778 FIVSQAVRSGSVSPTHYNVIYDNSGLK 804
                 | | : | :|    :    :|||
Db       725 LKSSTHVLAMSATPIPRTLQMSMTGLK 751


RESULT 12
CYAA_SACKL
ID   CYAA_SACKL     STANDARD;      PRT;  1839 AA.
AC   P23466;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Adenylate cyclase (EC 4.6.1.1) (ATP pyrophosphate-lyase) (Adenylyl
DE   cyclase).
GN   CYR1.
OS   Saccharomyces kluyveri (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4934;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91323718; PubMed=1864503;
RA   Young D., O'Neill K., Broek D., Wigler M.;
RT   "The adenylyl cyclase-encoding gene from Saccharomyces kluyveri.";
RL   Gene 102:129-132(1991).
CC   -!- FUNCTION: PLAYS ESSENTIAL ROLES IN REGULATION OF CELLULAR
CC       METABOLISM BY CATALYSING THE SYNTHESIS OF A SECOND MESSENGER,
CC       CAMP.
CC   -!- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC   -!- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-3 FAMILY.
CC   -!- SIMILARITY: CONTAINS 21 LEUCINE-RICH REPEATS (LRR).
CC   -!- SIMILARITY: CONTAINS 1 PP2C-LIKE DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X56042; CAA39513.1; -.
DR   PIR; S14464; OYBYK.
DR   PIR; JQ1145; JQ1145.
DR   InterPro; IPR001054; G_cyclase.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   InterPro; IPR001932; PP2C-like.
DR   InterPro; IPR000159; RA_domain.
DR   Pfam; PF00211; guanylate_cyc; 1.
DR   Pfam; PF00481; PP2C; 1.
DR   Pfam; PF00560; LRR; 13.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00044; CYCc; 1.
DR   SMART; SM00370; LRR; 4.
DR   SMART; SM00369; LRR_TYP; 3.
DR   SMART; SM00332; PP2Cc; 1.
DR   SMART; SM00314; RA; 1.
DR   PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
KW   Lyase; Repeat; Leucine-rich repeat; cAMP synthesis; Magnesium.
FT   REPEAT      632    655       LRR 1.
FT   REPEAT      659    679       LRR 2.
FT   REPEAT      680    702       LRR 3.
FT   REPEAT      703    724       LRR 4.
FT   REPEAT      726    748       LRR 5.
FT   REPEAT      749    771       LRR 6.
FT   REPEAT      773    793       LRR 7.
```

```
FT   REPEAT      794    818       LRR 8.
FT   REPEAT      820    833       LRR 9.
FT   REPEAT      834    857       LRR 10.
FT   REPEAT      859    879       LRR 11.
FT   REPEAT      880    903       LRR 12.
FT   REPEAT      905    926       LRR 13.
FT   REPEAT      927    949       LRR 14.
FT   REPEAT      951    974       LRR 15.
FT   REPEAT     1004   1025       LRR 16.
FT   REPEAT     1026   1048       LRR 17.
FT   REPEAT     1050   1072       LRR 18.
FT   REPEAT     1074   1096       LRR 19.
FT   REPEAT     1101   1124       LRR 20.
FT   REPEAT     1135   1160       LRR 21.
FT   DOMAIN     1185   1440       PP2C-LIKE.
FT   DOMAIN     1441   1839       CATALYTIC.
SQ   SEQUENCE   1839 AA;  206895 MW;  86A69BCB1F2733CB CRC64;

  Query Match             2.6%;  Score 116;  DB 1;  Length 1839;
  Best Local Similarity   18.8%;  Pred. No. 3.8;
  Matches  118;  Conservative   90;  Mismatches  206;  Indels  212;  Gaps   29;

Qy         41 PAEGELFGRGRQRGTAGGTAKSQGLQISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHV 100
              ||| | : | |   |:||   :| |    |  |:|      |   | ||       | |:
Db       1327 PAEREEYERIR---TSGGYVNNQKLD---GVSEVS------RAVGFFDL-------LPHI 1367

Qy        101 KESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDY------------NPLMEARRLR-- 146
              |    |  ::  |     : :   |     ::||             |: | :::
Db       1368 HAS-PDISETVLSYSDEMLIIATHKLWE----YLDYETVCDISRENKSQPMSAAEKMKDY 1422

Qy        147 ----------------------SALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTE-- 180
                                   |     |||| : : |: |:|   :||| ::
Db       1423 AISYGCSDNITILCVSLDKSVNQQSQFTLNREDLISRKNTFEDTVL---RRLQPEIAPPT 1479

Qy        181 -----VFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNP 235
                  ||:      |:: :  |    |       :: :| | || |:|      |
Db       1480 GNVAIVFT------DIKNSTFLWELFPDAMRAAIKTHNDIMRRQLRIYG----GYEVKTE 1529

Qy        236 NDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKF 295
              |       :| :|  |       :::: |  |  |:| :|                 :
Db       1530 GDAF------MVAFPTPT-------SALVWCLSVQLKLLEAE----------------W 1559

Qy        296 QEQVSKEXIGLVVLTKYNIKTY-------------RVDDID-----WDQNPKSTFKKADG 337
              |:::    | ::      | |               | :||      |       | |
Db       1560 PEEITSIQDGCLITDNSGTKVYLGLSVRMGVHWGCPVPEIDLVTQRMDYLGPVVNKAARV 1619

Qy        338 SGVS----------FLEYYRK--QYNQEITDLKQPVLVSQPKRRRGP---GGTLPGPAML 382
              |||:          |   ::|  :::: : : ::|:     |   |    |  |     :
Db       1620 SGVADGGQITLSSDFCSEFKKIMKFHKRVVENQEPL-----KEVYGEDFIGEVLEREIHM 1674

Qy        383 IPELCY---------LTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDN 433
              :  : :         | ||  |         | ||    |  : :   |         ||:
Db       1675 LENVGWVFKDLGEQKLKGLETKEFITIAYPKTLASRHDLATKNQNSSV--------LNDD 1726

Qy        434 VQRELRDWGLSFDSNLLSFSGRILQTEKIHQGG--KTFDYNPQFADWSKETRGAPLISVK 491
              :  :||      :: | | :| ::::|         ||| | : |  :| |
Db       1727 LLFQLRTISNKLENILSSINGGLIESETPGNSSIYMTFDKNTKDAVITKSTE-------- 1778

Qy        492 PLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDRTEAYLRVLQ--QKV 549
              :|:              | : :|  ||                  | |: : :||  ||:
Db       1779 --SDWI-----------SFLDHL--VT----------------RVESTVAILQLRQKL 1805

Qy        550 TADTQIVVCLLSSNRKDKYDAIKKYL 575
                ::     |:  |  :: : : |
Db       1806 QGGLELYTSSDSTMHKSVFELLDEIL 1831


RESULT 13
HMP2_YEREN
ID   HMP2_YEREN     STANDARD;      PRT;  2035 AA.
AC   P48633;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   High-molecular-weight protein 2 (HMWP2).
GN   IRP2.
OS   Yersinia enterocolitica.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Yersinia.
OX   NCBI_TaxID=630;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=8081 / Serotype O:8;
RX   MEDLINE=93374844; PubMed=8366034;
RA   Guilvout I., Mercereau-Puijalon O., Bonnefoy S., Pugsley A.P.,
RA   Carniel E.;
RT   "High-molecular-weight protein 2 of Yersinia enterocolitica is
RT   homologous to AngR of Vibrio anguillarum and belongs to a family of
RT   proteins involved in nonribosomal peptide synthesis.";
RL   J. Bacteriol. 175:5488-5504(1993).
CC   -!- FUNCTION: UNKNOWN. MAY BE INVOLVED IN THE NONRIBOSOMAL SYNTHESIS
CC       OF SMALL PEPTIDES.
CC   -!- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- DOMAIN: CONSISTS OF A CENTRAL REGION WITH SIMILARITY TO THE REPEAT
CC       DOMAINS OF ACVS AND GRC2, FLANKED BY TWO REPEAT DOMAINS, EACH OF
CC       WHICH CONTAINS 5 DIRECT REPEATS.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L18881; AAA27636.1; -.
DR   EMBL; Z35454; CAA84606.1; -.
DR   PIR; A48654; A48654.
DR   HSSP; P14687; 1AMU.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; Condensatn.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF00501; AMP-binding; 1.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00668; Condensation; 2.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00455; AMP_BINDING; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Ligase; Multifunctional enzyme; Phosphopantetheine; Repeat.
FT   DOMAIN        3    547       I.
FT   REPEAT      114    146       I-DR1.
FT   REPEAT      310    321       I-DR2.
FT   REPEAT      378    390       I-DR3.
FT   REPEAT      454    462       I-DR4.
FT   REPEAT      477    491       I-DR5.
FT   DOMAIN     1466   1919       II.
FT   REPEAT     1495   1527       II-DR1.
FT   REPEAT     1682   1693       II-DR2.
FT   REPEAT     1750   1762       II-DR3.
FT   REPEAT     1826   1834       II-DR4.
FT   REPEAT     1849   1863       II-DR5.
FT   DOMAIN       20     88       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1409   1475       ACYL CARRIER (ACP) 2.
FT   DOMAIN     1944   2014       ACYL CARRIER (ACP) 3.
FT   BINDING      52     52       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   2035 AA;  228826 MW;  1C801377A4375BDC CRC64;

  Query Match             2.5%;  Score 115.5;  DB 1;  Length 2035;
  Best Local Similarity   18.4%;  Pred. No. 4.9;
  Matches  170;  Conservative  133;  Mismatches  365;  Indels  255;  Gaps   43;
```

```
Qy          11 GRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRGTAGGTAKSQGLQISAG 70
Db        1060 GHTAGADPIPLMNSLA------IQPRWQAVVERWLAFLVTQRRLKPA----AEGYQVCAG 1109

Qy          71 FQELSXAERGGRRRDF--HDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWA 128
Db        1110 ------EEREDEHPHFSGHDLTLSQ----------------ILRGARNELSLLNDAQWS 1146

Qy         129 LYQYHIDYNPLMEARRLRSALLFQH-------EDLIGKCHAFDGTILFLPKRLQQKVTEV 181
Db        1147 ----------------PESLAFNHPASAPYIQELATICQQ-------LAQRLQRPVRLL 1182

Qy         182 FSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDI 241
Db        1183 EVGTRTGRAAESLLAQLN----------------AGQIEYVGLEQ-------SQEMLLS 1218

Qy         242 PSHRLVIWPGF--------TTSILQYENSIMLCTDVSHKVLRSE--------------TV 279
Db        1219 ARQRLASWPGARLSPWNADTLAAHAHSGDIIWLNNALHRLLPEDPGLLATLQQLAVPGAL 1278

Qy         280 LDFMFNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSG 339
Db        1279 LYVM----------EFRQLTPSALLSTLLLTNGQPEALLHNSADW----AALFSAA---- 1320

Qy         340 VSFLEYYRKQYNQEITDLKQPVLVSQPKR--RRGP-------GGTLPGPAMLIPE-LCYL 389
Db        1321 -----AFNCQHSDEVAGL-QRFLVQCPDRQVRRDPRQLQAALAGRLPG--WMVPQRIVFL 1372

Qy         390 TG--LTDKMRNDFNVMKDLAVHTRLTPEQR-----------QREVGRLIDYIHKNDNVQ 435
Db        1373 DALPLTANGKIDYQALK-----RRHTPKAENQAEADLPQGDIEKQVAALWQQLLSTGNVT 1427

Qy         436 RELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISV----- 490
Db        1428 RETDFFQQGGDSLLAT---RL--TGQLHQAG----YEAQLSDLFNHPRLADFAATLRKID 1478

Qy         491 KPLDNWLLIYTRRNYEA-ANSLIQNLFKVTPAMGMQMRKA-----IMIEVDDRTEAYLRV 544
Db        1479 VPVEQPFVHSPEERYQPFALTDVQQAYLVGRQPGFTLGGVGSHFFVEFEIADLDLTRLET 1538

Qy         545 LQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKI 604
Db        1539 VWNRLIARHDMLRAVVLDGQQQVLE----------QTPPWVIPTHTLHTPEEALRVREKL 1588

Qy         605 ALQMNCKMGGELWRV-DIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCI 663
Db        1589 AHQV---LNPEVWPVFDLQV----------------------GYV----DGMPARLWLCL 1619

Qy         664 FQDRGQELVDGL--KVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFL 721
Db        1620 ----DNLLLDGLSMQILLAELEHGYRYPQQLLPPLPVTFRDYLQQPSLQSPNPDSLAWWQ 1675

Qy         722 DCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNP----LPGTVIDVEVTRPEWYDF 777
Db        1676 AQLDDIPPA--PALPLRCLPQEVETPRFARLNGALDSTRWHRLKKRAADAHLT-PSAVLL 1732

Qy         778 FIVSQAVRSGSVSP--THYNVIYDNSGLKPDHIQRLTYKLCHIYYNW-PGVIRVPAPCQY 834
Db        1733 SVWSTVLSAWSAQPEFTLNLTLFDRRPLHPQINQILGDFTSLMLLSWHPGE-------SW 1785

Qy         835 AHKLAFLVGQSIHREPNLSLSNR 857
Db        1786 LHS-----AQSLQQRLSQNLNHR 1803
```

```
RESULT 14
CLH_YEAST
ID   CLH_YEAST      STANDARD;      PRT;  1653 AA.
AC   P22137;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Clathrin heavy chain.
GN   CHC1 OR YGL206C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=91093343; PubMed=1898742;
RA   Lemmon S.K., Pellicena-Palle A., Conley K., Freund C.L.;
RT   "Sequence of the clathrin heavy chain from Saccharomyces cerevisiae
RT   and requirement of the COOH terminus for clathrin function.";
RL   J. Cell Biol. 112:65-80(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97298309; PubMed=9153757;
RA   Feuermann M., Simeonava L., Souciet J.-L., Potier S.;
RT   "Analysis of 21.7 kb DNA sequence from the left arm of chromosome VII
RT   reveals 11 open reading frames: two correspond to new genes.";
RL   Yeast 13:475-477(1997).
CC   -!- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF
CC       COATED PITS & VESICLES. IN YEAST, IT IS INVOLVED IN THE RETENTION
CC       OF PROTEINS IN AN INTRACELLULAR MEMBRANE COMPARTMENT, PRESUMABLY
CC       THE TRANS-GOLGI.
CC   -!- SUBUNIT: CLATHRIN TRISKELIONS, COMPOSED OF 3 HEAVY CHAINS AND 3
CC       LIGHT CHAINS, ARE THE BASIC SUBUNITS OF THE CLATHRIN COAT.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND
CC       VESICLES.
CC   -!- DOMAIN: THE C-TERMINAL THIRD OF THE HEAVY CHAINS FORMS THE HUB OF
CC       THE TRISKELION. THIS REGION CONTAINS THE TRIMERIZATION DOMAIN AND
CC       THE LIGHT-CHAIN BINDING DOMAIN INVOLVED IN THE ASSEMBLY OF THE
CC       CLATHRIN LATTICE.
CC   -!- SIMILARITY: BELONGS TO THE CLATHRIN HEAVY CHAIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52900; CAA37082.1; -.
DR   EMBL; Z72728; CAA96919.1; -.
DR   PIR; A36349; A36349.
DR   HSSP; P11442; 1BPO.
DR   SGD; S0003174; CHC1.
DR   InterPro; IPR001473; Clathrin_propel.
DR   InterPro; IPR000547; Clathrin_repeat.
DR   Pfam; PF00637; Clathrin; 7.
DR   Pfam; PF01394; Clathrin_propel; 3.
DR   SMART; SM00299; CLH; 7.
KW   Coated pits.
FT   DOMAIN        1    483       GLOBULAR TERMINAL DOMAIN.
FT   DOMAIN      484    527       FLEXIBLE LINKER.
FT   DOMAIN      528   1653       HEAVY CHAIN ARM.
FT   DOMAIN      528      ?       DISTAL SEGMENT.
FT   DOMAIN        ?   1653       PROXIMAL SEGMENT.
FT   DOMAIN      453    469       BINDING SITE FOR THE UNCOATING ATPASE,
FT                                INVOLVED IN LATTICE DISASSEMBLY
FT                                (POTENTIAL).
FT   BINDING    1219   1528       LIGHT CHAIN (BY SIMILARITY).
SQ   SEQUENCE   1653 AA;  187233 MW;  622F3083DF2FE315 CRC64;

  Query Match             2.5%;  Score 114;  DB 1;  Length 1653;
  Best Local Similarity   18.9%;  Pred. No. 4.6;
  Matches  143;  Conservative  110;  Mismatches  281;  Indels  222;  Gaps   34;

Qy         170 LPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLK--IMNLQQ 227
Db         245 LPSQYQKETTDIFFPPDATNDFPIAVQVSEK----------YGIIY-LLTKYGFIHLYE 292
```

```
Qy        228 I--GRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSE----TVLD 281
Db        293 LETGTNLF---------VNRITAESVFTAAPYNHENGI-ACINKKGQVLAVEISTSQIVP 342

Qy        282 FMFNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVS 341
Db        343 YILN-----------KLSNVALALIVATRGGLP--GADDL------------------- 369

Qy        342 FLEYYRKQYNQEI---------------TDLKQPVLVSQPKRRRGPGGTLPGPAMLIPEL 386
Db        370 ----FQKQFESLLLQNDYQNAAKVAASSTSLRNQNTINRLKNIQAP------PGAISPIL 419

Qy        387 CYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQ--RELRDWGLS 444
Db        420 LYFSTLLDKGK----LNKEETIELARPVLQQDRK--QLFEKWLKEDKLECSEELGDIVKP 473

Qy        445 FDSNL--------------LSFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISV 490
Db        474 FDTTLALACYLRAGAHAKVISCLAELQQFEKIIPYCQKVGYQPNF-----------LVLI 522

Qy        491 KPLDNWLLIYTRRNYEAANSLIQN--------LFKVTPAMGMQ---------MRKAIMIE 533
Db        523 SSL---IRSSPDRASEFAVSLLQNPETASQIDIEKIADLFFSQNHIQQGTSLLLDALKGD 579

Qy        534 VDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYD----------------AIKKYLCT 577
Db        580 TPDQGHLQTRVLEVNLLHAPQVADAILGNNIFSHYDKPTIASLSEKAGLYQRALENY--T 637

Qy        578 DCPTPSQCVVART----------LGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVM 627
Db        638 DIKDIKRCVVHTNALPIDWLVGYFGKLNVEQSLACLKALMDNNIQANIQTVVQVATKFSD 697

Qy        628 IVG----IDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAAL 683
Db        698 LIGPSTLIKLFEDYNA-----------TEGLYYYLASLVNLTEDKDVV--YKYIEAAAK 743

Qy        684 --------RAWNSCNEYMPSRIIVY-RDGVGDGQLKTLVNYE-------------VPQFL 721
Db        744 MKQYREIERIVKDNNVYDPERVKNFLKDANLEDQLPLVIVCDRFDFVHEMILYLYKSQNL 803

Qy        722 DCLKSIGRGYNPRXTVIVVKKRVN---TRFFAQSGGRLQNPL---PGTVIDVEVTRPEWY 775
Db        804 KFIETYVQQVNPSKTAQVVGALLDMDCDEAFIQS--LLQSVLGQVPINELTTEVEKRNRL 861

Qy        776 DFFI--VSQAVRSGSVSPTHYNV---IYDNSGLKPD 806
Db        862 KILLPFLEQSLSQGIQDQAVYNALAKIYIDSNNSPE 897
```

```
RESULT 15
YMW6_YEAST
ID   YMW6_YEAST     STANDARD;      PRT;  1277 AA.
AC   Q04264; Q04780;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hypothetical 147.0 kDa protein in ABF2-CHL12 intergenic region.
GN   YMR076C OR YM9582.01C OR YM9916.15C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE OF 1-682 FROM N.A.
RC   STRAIN=S288c / AB972;
RA   Gentles S., Bowman S., Barrell B.G., Rajandream M.A.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 659-1227 FROM N.A.
RC   STRAIN=S288c / AB972;
RA   Pearson D., Bowman S., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z49259; CAA89222.1; -.
DR   EMBL; Z48952; CAA88801.1; -.
DR   SGD; S0004681; YMR076C.
KW   Hypothetical protein.
SQ   SEQUENCE   1277 AA;  147040 MW;  9DF40A5274FD9623 CRC64;
```

```
  Query Match             2.4%;  Score 111;  DB 1;  Length 1277;
  Best Local Similarity   19.4%;  Pred. No. 5.2;
  Matches   79;  Conservative   69;  Mismatches  155;  Indels  104;  Gaps   18;

Qy         97 LDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQH--- 153
Db        841 LSDTKYSAIGNKVFTLKLFTNKLRSIAP--------DVPRDELAESFTEKTMKLFFYLIA 892

Qy        154 --EDLIGKCHAFDGTILFLPKRLQQK--------------VTEVFSKTRNGEDVRITITL 197
Db        893 SGGELISE---FNKEFYPTPSNYQTKLRCVAGIQVLKLARISNLNNFIKPSDIIKLINLV 949

Qy        198 TNELPPTSPTCL-QFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSI 256
Db        950 EDESLPVRKTFLEQLKDYVANELISIKFLPLVFFTAYEPDVELKTTTK---IWINFTFGL 1006

Qy        257 LQYENSIM-------LCTDVSHKV-----LRSE---------TVLDFMFNFYHQTEEHKF 295
Db       1007 KSFKKGTIFERALPRLIHAIAHHPDIVGGLDSEGDAYLNALTTAIDYLL-FY-------F 1058

Qy        296 QEQVSKEXIGLVVLTKYNIKTYR----VDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYN 351
Db       1059 DSIAAQENFSLLYYLSERVKNYQDKLVEDEIDEEEGPQK-----EEAPKKHRPYGQKMY- 1112

Qy        352 QEITDLKQPVLVSQPKRRRGPGGTLPGPAML--------------------IPELCYLT 390
Db       1113 -IIGELSQMILLNLKEKKNWQHSAYPGKLNLPSDLFKPFATVQEAQLSFKTYIPE----- 1166

Qy        391 GLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRE 437
Db       1167 SLTEKIQNNIKAKIGRILHTSQT--QRQRLQKRLL--AHENNESQKK 1209
```

Search completed: July 22, 2003, 07:03:03
Job time : 13.0663 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 06:19:03 ; Search time 33.5542 Seconds
(without alignments)
5287.169 Million cell updates/sec

Title: US-09-873-737A-6
Perfect score: 4534
Sequence: 1 MTGRARARARGRARGQETAQ.....VGQSIHREPNLSLSNRLYYL 861

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4528	99.9	861	4	O95404	O95404 homo sapien
2	4513	99.5	861	4	Q96J94	Q96j94 homo sapien
3	4498	99.2	861	4	Q8TBY5	Q8tby5 homo sapien
4	4314.5	95.2	862	11	Q9JMB7	Q9jmb7 mus musculu
5	4061	89.6	775	4	Q96JD5	Q96jd5 homo sapien
6	3026.5	66.8	858	13	Q8UVX0	Q8uvx0 brachydanio
7	2033	44.8	854	5	Q9GPA8	Q9gpa8 strongyloce
8	1835	40.5	808	5	Q9GPA7	Q9gpa7 strongyloce
9	1742	38.4	971	11	Q9JMB6	Q9jmb6 mus musculu
10	1730	38.2	973	4	Q8TC59	Q8tc59 homo sapien
11	1464	32.3	866	5	O76922	O76922 drosophila
12	1417	31.3	843	5	Q9VKM1	Q9vkm1 drosophila
13	1365.5	30.1	580	11	Q99MV6	Q99mv6 mus musculu
14	1291.5	28.5	824	5	P90786	P90786 caenorhabdi
15	1220	26.9	530	4	Q96SW6	Q96sw6 homo sapien
16	1204	26.6	421	5	Q9GPA6	Q9gpa6 strongyloce
17	1144	25.2	722	5	Q17567	Q17567 caenorhabdi
18	959	21.2	371	4	Q9NW28	Q9nw28 homo sapien
19	718	15.8	781	5	Q9U5C9	Q9u5c9 paramecium
20	533	11.8	1014	10	Q9SHF3	Q9shf3 arabidopsis
21	518	11.4	1194	10	Q9SHF2	Q9shf2 arabidopsis
22	500.5	11.0	891	5	O16720	O16720 caenorhabdi
23	498.5	11.0	1123	10	Q9LP83	Q9lp83 arabidopsis
24	493	10.9	950	5	Q9V3L2	Q9v3l2 drosophila
25	492	10.9	1000	5	Q20578	Q20578 caenorhabdi
26	490	10.8	984	5	Q9V6V6	Q9v6v6 drosophila
27	486	10.7	860	4	Q9H9G7	Q9h9g7 homo sapien
28	485.5	10.7	924	10	Q9ZVD5	Q9zvd5 arabidopsis
29	484	10.7	924	4	Q9HCK5	Q9hck5 homo sapien
30	475	10.5	851	4	Q8TCZ5	Q8tcz5 homo sapien
31	470.5	10.4	990	10	Q9C793	Q9c793 arabidopsis
32	461	10.2	904	10	Q9SDG8	Q9sdg8 oryza sativ
33	432	9.5	640	4	Q9H1U6	Q9h1u6 homo sapien
34	430.5	9.5	922	3	Q96UM0	Q96um0 blumeria gr
35	430.5	9.5	1035	5	Q21079	Q21079 caenorhabdi
36	422	9.3	887	10	O48771	O48771 arabidopsis
37	421	9.3	962	5	Q9VUQ6	Q9vuq6 drosophila
38	421	9.3	1037	5	Q9TW94	Q9tw94 caenorhabdi
39	420	9.3	601	5	Q960T1	Q960t1 drosophila
40	417.5	9.2	585	4	Q8WV58	Q8wv58 homo sapien
41	388	8.6	898	5	Q9XVI3	Q9xvi3 caenorhabdi
42	379	8.4	377	4	Q96ID1	Q96id1 homo sapien
43	379	8.4	377	11	Q8R3Q7	Q8r3q7 mus musculu
44	377	8.3	694	5	Q8T3N2	Q8t3n2 drosophila
45	373.5	8.2	424	5	Q8T059	Q8t059 drosophila

ALIGNMENTS

RESULT 1
O95404
ID O95404 PRELIMINARY; PRT; 861 AA.
AC O95404;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE HIWI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE=99069219; PubMed=9851978;
RA Cox D.N., Chao A., Baker J., Chang L., Qiao D., Lin H.;
RT "A novel class of evolutionarily conserved genes defined by piwi are
RT essential for stem cell self-renewal.";
RL Genes Dev. 12:3715-3727(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Qiao D., Zeeman A.-M., Deng W., Looijenga L.H.J., Lin H.;
RT "Molecular characterization of hiwi, a human member of the piwi stem
RT cell gene family whose overexpression is correlated to seminomas.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF104260; AAC97371.2; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
SQ SEQUENCE 861 AA; 98530 MW; 43D7F60E99D997B7 CRC64;

Query Match 99.9%; Score 4528; DB 4; Length 861;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 858; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRGTAGGTA 60

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRGTAGGTA 60

Qy         61 KSQGLQISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFR 120
                 ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 KSQGLQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFR 120

Qy        121 LTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTE 180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTE 180

Qy        181 VFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPID 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPID 240

Qy        241 IPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVS 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVS 300

Qy        301 KEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQP 360
                 || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KELIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQP 360

Qy        361 VLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQRE 420
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQRE 420

Qy        421 VGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSK 480
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 VGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSK 480

Qy        481 ETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDRTEA 540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDRTEA 540

Qy        541 YLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAI 600
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAI 600

Qy        601 ATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFS 660
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFS 660

Qy        661 RCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQF 720
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQF 720

Qy        721 LDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIV 780
                 |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 LDCLKSIGRGYNPRLTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIV 780

Qy        781 SQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAF 840
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 SQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAF 840

Qy        841 LVGQSIHREPNLSLSNRLYYL 861
                 |||||||||||||||||||
Db        841 LVGQSIHREPNLSLSNRLYYL 861


RESULT 2
Q96J94
ID   Q96J94      PRELIMINARY;      PRT;   861 AA.
AC   Q96J94;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   PIWI protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Sha J.H.;
RT   "Cloning and identification of human piwi protein related to testis
RT   development.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF387507; AAK69348.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   861 AA;  98603 MW;  58D7F6C7321DEFA4 CRC64;

  Query Match             99.5%;  Score 4513;  DB 4;  Length 861;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches  856;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRGTAGGTA 60
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRGTAGGTA 60

Qy         61 KSQGLQISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFR 120
                 ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 KSQGLQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFR 120

Qy        121 LTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTE 180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTE 180

Qy        181 VFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPID 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPID 240

Qy        241 IPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVS 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVS 300

Qy        301 KEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQP 360
                 || ||||||||||| |||||||||||||||||||||||| ||||||||||||||||||||||||
Db        301 KELIGLVVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQP 360

Qy        361 VLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQRE 420
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQRE 420

Qy        421 VGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSK 480
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 VGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSK 480

Qy        481 ETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDRTEA 540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDRTEA 540

Qy        541 YLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAI 600
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAI 600

Qy        601 ATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFS 660
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFS 660

Qy        661 RCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQF 720
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQF 720

Qy        721 LDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIV 780
                 |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 LDCLKSIGRGYNPRLTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIV 780

Qy        781 SQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAF 840
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 SQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAF 840
```

```
Qy        841 LVGQSIHREPNLSLSNRLYYL 861
              |||||||||||||||||||||
Db        841 LVGQSIHREPNLSLSNRLYYL 861


RESULT 3
Q8TBY5
ID   Q8TBY5      PRELIMINARY;      PRT;   861 AA.
AC   Q8TBY5;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Piwi-like 1 (Drosophila).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Strausberg R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC028581; AAH28581.1; -.
SQ   SEQUENCE   861 AA;  98545 MW;  D33376EEDEDF743A CRC64;

  Query Match             99.2%;  Score 4498;  DB 4;  Length 861;
  Best Local Similarity   99.1%;  Pred. No. 0;
  Matches  853;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRGTAGGTA 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRGTAGGTA 60

Qy         61 KSQGLQISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFR 120
              ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db         61 KSQGLQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFR 120

Qy        121 LTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTE 180

Qy        181 VFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPID 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPID 240

Qy        241 IPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVS 300

Qy        301 KEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQP 360
              || |||||||||| |||||||||||||||||||||| ||||||||||||||||||||||||
Db        301 KELIGLVVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQP 360

Qy        361 VLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQRE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQRE 420

Qy        421 VGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 VGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSK 480

Qy        481 ETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDRTEA 540
              |||||||||| |||||||||||||||||||||||||||||:|||||||||||||||||||
Db        481 ETRGAPLISVNPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMKKAIMIEVDDRTEA 540

Qy        541 YLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAI 600
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        541 YLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYPCTDCPTPSQCVVARTLGKQQTVMAI 600

Qy        601 ATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFS 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFS 660

Qy        661 RCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQF 720

Qy        721 LDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIV 780
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        721 LDCLKSIGRGYNPRLTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIV 780

Qy        781 SQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 SQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAF 840

Qy        841 LVGQSIHREPNLSLSNRLYYL 861
              |||||||||||||||||||||
Db        841 LVGQSIHREPNLSLSNRLYYL 861


RESULT 4
Q9JMB7
ID   Q9JMB7      PRELIMINARY;      PRT;   862 AA.
AC   Q9JMB7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   MIWI (piwi).
GN   PIWIL1 OR MIWI.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Miyagawa S.K., Kimura T., Nakano T.;
RT   "Molecular Cloning and characterization of Piwi family genes.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Deng W., Lin H.;
RT   "Miwi, a murine homolog of piwi, encodes a cytoplasmic protein
RT   essential for spermatogenesis.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB032604; BAA93705.1; -.
DR   EMBL; AF438405; AAL31014.1; -.
DR   MGD; MGI:1928897; Piwil1.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   862 AA;  98574 MW;  45588D13284CCC4C CRC64;

  Query Match             95.2%;  Score 4314.5;  DB 11;  Length 862;
  Best Local Similarity   94.3%;  Pred. No. 0;
  Matches  813;  Conservative   22;  Mismatches   26;  Indels    1;  Gaps    1;

Qy          1 MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRG-TAGGT 59
              |||||||||||||||||| | ||: |||||||| ||||  | ||:| |||||||   | |
Db          1 MTGRARARARGRARGQETVQHVGAAASQQPGYIPPRPQQSPTEGDLVGRGRQRGMVVGAT 60

Qy         60 AKSQGLQISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHF 119
              :||| |||||||||| ||||||||||||||||||||||||||||||:||||||
Db         61 SKSQELQISAGFQELSLAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVKLSTNHF 120

Qy        120 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVT 179
              ||||||||||||||||||||||||||||||||||||:|||||||||||||||| |||
Db        121 RLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGRCHAFDGTILFLPKRLQHKVT 180

Qy        180 EVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPI 239
              ||||:||||| |||||||||||||||||||||||||||||||||||||||||:|||
Db        181 EVFSQTRNGEHVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPSDPI 240

Qy        240 DIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQV 299
              |||:||||||||||||||:|||||||||||||||||||| | ||||||||||
```

```
Db        241 DIPNHRLVIWPGFTTSILQYENNIMLCTDVSHKVLRSETVLDFMFNLYQQTEEHKFQEQV 300

Qy        300 SKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQ 359
              ||| |||:|||||| ||||||||||||||||||||| ||||||||||||||||||||||
Db        301 SKELIGLIVLTKYNNKTYRVDDIDWDQNPKSTFKKADGSEVSFLEYYRKQYNQEITDLKQ 360

Qy        360 PVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 PVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQR 420

Qy        420 EVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWS 479
              ||||||||||:||||||||||||||||||||||||:||||||||||||||||||||||||
Db        421 EVGRLIDYIHKDDNVQRELRDWGLSFDSNLLSFSGRILQSEKIHQGGKTFDYNPQFADWS 480

Qy        480 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDRTE 539
              ||||||||||||||||||||||||||||||||||||||||||:||:||||||||||||||
Db        481 KETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGIQMKKAIMIEVDDRTE 540

Qy        540 AYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 599
              |||| |||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AYLRALQQKVTSDTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMA 600

Qy        600 IATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWF 659
              ||||||||||||||||||||:|||| |||||||||| ||||||||||||||||||||||
Db        601 IATKIALQMNCKMGGELWRVDMPLKLAMIVGIDCYHDTTAGRRSIAGFVASINEGMTRWF 660

Qy        660 SRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQ 719
              |||:|||||||||||||||||||||: ||||||||:||||||||||||||||||||||||
Db        661 SRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDGQLKTLVNYEVPQ 720

Qy        720 FLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 779
              |||||||:||||||| |||||||||| ||||||||||||||||||||||||||||||||
Db        721 FLDCLKSVGRGYNPRLTVIVVKKRVNARFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 780

Qy        780 VSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLA 839
              ||||||||||||||||||:|||||||||||||||:||||||||||||||||||||||||||
Db        781 VSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKLCHVYYNWPGVIRVPAPCQYAHKLA 840

Qy        840 FLVGQSIHREPNLSLSNRLYYL 861
              ||||||||||||||||||||||
Db        841 FLVGQSIHREPNLSLSNRLYYL 862
```

RESULT 5
Q96JD5
ID Q96JD5 PRELIMINARY; PRT; 775 AA.
AC Q96JD5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE HIWI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Sharma A.K., Nelson M.C., Brandt J.E., Wessman M., Muhmud N.,
RA Weller K.P., Hoffman R.;
RT "Human CD34+ Stem Cells Express the hiwi Gene, a Human Homolog of the
RT Drosophila Gene piwi.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF264004; AAK92281.1; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
SQ SEQUENCE 775 AA; 89484 MW; DF169A2E9EAFD916 CRC64;

Query Match 89.6%; Score 4061; DB 4; Length 775;
Best Local Similarity 99.5%; Pred. No. 2.6e-303;
Matches 768; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy         90 GVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSAL 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 GVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSAL 63

Qy        150 LFQHEDLIGKCHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCL 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 LFQHEDLIGKCHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCL 123

Qy        210 QFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDV 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 QFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDV 183

Qy        270 SHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPK 329
              |||||||||||||||||||||||||||||| |||||||| |||||||||||||||||||
Db        184 SHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKELIGLVVLTKYNNKTYRVDDIDWDQNPK 243

Qy        330 STFKKADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYL 389
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 STFKKADGSEVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYL 303

Qy        390 TGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNL 449
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 TGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNL 363

Qy        450 LSFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAAN 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 LSFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAAN 423

Qy        510 SLIQNLFKVTPAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYD 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 SLIQNLFKVTPAMGMQMRKAIMIEVDDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYD 483

Qy        570 AIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIV 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 AIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIV 543

Qy        630 GIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSC 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 GIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSC 603

Qy        690 NEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFF 749
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        604 NEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVVKKRVNTRFF 663

Qy        750 AQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQ 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 AQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQ 723

Qy        810 RLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 861
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 RLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 775
```

RESULT 6
Q8UVX0
ID Q8UVX0 PRELIMINARY; PRT; 858 AA.
AC Q8UVX0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Piwi protein.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Weeraratne S.D., Gong Z., Tan C.-H.;
RT "Cloning and characterization of zebrafish homolog of piwi, essential
RT for germ-line stem cell self-renewal.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.

```
DR   EMBL; AF336369; AAL57170.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   858 AA;  97451 MW;  6A12F2E511465777 CRC64;

  Query Match             66.8%;  Score 3026.5;  DB 13;  Length 858;
  Best Local Similarity   64.1%;  Pred. No. 9.9e-224;
  Matches  554;  Conservative  130;  Mismatches  171;  Indels    9;  Gaps    3;

Qy          1 MTGRARARARGRARGQETAQLVGSTASQQPGYIQPRPQP---PPAEGELFGRGRQRGTAG 57
              ||||||||:||| |||| |          ||   |   :|    |:||:| |||||:   |
Db          1 MTGRARARSRGRGRGQEPA-----APGAQPPVSQEAAKPVVSTPSEGQLVGRGRQKPAPG 55

Qy         58 GTAKSQGLQISAGFQELSXAERGGRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTN 117
                ::   ||||||||::   ||||||||||| ||:||| ::||||||| ||  : |  |
Db         56 AMSEEAMLQISAGFQQVKIGERGGRRRDFHDSGVHTRQLMEHVKESKTGVSGTAIELRAN 115

Qy        118 HFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQK 177
              || ||| |||||||:|| | ||:|||||||||||: :|| | ||| ||||| :|:
Db        116 FMRLLSRPMWALYQYHVDYKPPMESRRLRSALLFQHEETLGKAHTFDGAILFLPNKLRNA 175

Qy        178 VTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPND 237
              | : |:||||| | ||:||||||||:|| |||||||:|||:|:|:|:|||||:||||:|
Db        176 ETVLCSETRNGEKVEITVTLTNELPPSSPVCLQFYNILFRRILRILNMQQIGRHYYNPDD 235

Qy        238 PIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQE 297
              | :|| ||| ||||| |:|||||:|||||:||||||||||||||::   |  : :| |
Db        236 PFNIPQHRLTIWPGFMTTILQYESSIMLCSDVSHKVLRSETVLDFMYSLRQQCGDQRFPE 295

Qy        298 QVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDL 357
              :|| :|||::|||| |||||:||| ||  | :|||| | : :||  |:: ||  :|||
Db        296 ACTKELVGLIILTKYNNKTYRIDDIAWDHTPNNTFKKGD-TEISFKNYFKSQYGLDITDG 354

Qy        358 KQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQR 417
              | :|||  ||    |  ||||||:|| |||||||||| |||:||||| ||||:|||| 
Db        355 NQVLLVSHVKRLGPSGRPPPGPAMLVPEFCYLTGLTDKMRADFNIMKDLASHTRLSPEQR 414

Qy        418 QREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFAD 477
              :  : |||  |::| :|| ||  |||||:: ||| :||:| :|:| |||: |:||| ||
Db        415 EGRINRLISNINRNGDVQNELTTWGLSFENKLLSLNGRVLPSERIIQGGRAFEYNPWTAD 474

Qy        478 WSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEVDDR 537
              |||| || ||||   |||||: ||||| : | ||:| | ||:  ||::|:::|:||| :||
Db        475 WSKEMRGLPLISCMSLDNWLMFYTRRNADVAQSLLQTLNKVSGPMGIRMQRAVMIEYEDR 534

Qy        538 TEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTV 597
              |: || ||| |  :||:|| :| :||||||| :||||| |||||||||:||: | | :
Db        535 QESLLRALQQNVARETQMVVVILPTNRKDKYDCVKKYLCVDCPTPSQCVVSRTISKPQAL 594

Qy        598 MAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTR 657
              | :|||||||||||||||| |:||: :|||||||||  ||:||  |||:|:||:| 
Db        595 MTVATKIALQMNCKMGGELWSVEIPLRQLMIVGIDCYHDTAAGKRSIGAMVASLNQGMSR 654

Qy        658 WFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEV 717
              |||:|: |:||||::| ||  || ||:|:   | :|||||||||||| |:::|:|||
Db        655 WFSKCVLQNRGQEIIDALKGSLQGALKAYLKYNNSLPSRIIVYRDGVGDGMLQSVVDYEV 714

Qy        718 PQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDF 777
              || :  :|::|: | |: :|:|||||:::|||||:  |:: || |||||| ||||||||| 
Db        715 PQIMQSIKTMGQDYEPKLSVVVVKKRISSRFFARIDGKIANPPPGTVIDTEVTRPEWYDF 774

Qy        778 FIVSQAVRSGSVSPTHYNVIYIDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHK 837
              ||||||| | |:|||||::|||||||:|||||||:|||| |::||||||||| 
Db        775 FIVSQAVRFGCVAPTHYNVVFDNSGLKPDHMQRLTYKLCHMYYNWQGIVRVPAPCQYAHK 834

Qy        838 LAFLVGQSIHREPNLSLSNRLYYL 861
              ||||||||||:|||::| : ||||
Db        835 LAFLVGQSIHKEPNMNLDDFLYYL 858


RESULT 7
```

```
Q9GPA8
ID   Q9GPA8      PRELIMINARY;      PRT;   854 AA.
AC   Q9GPA8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Seawi.
OS   Strongylocentrotus purpuratus (Purple sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Strongylocentrotus.
OX   NCBI_TaxID=7668;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rodriguez A.J., Bonder E.M.;
RT   "Seawi - Cloning and Molecular Characterization of a Sea Urchin
RT   Homolog of Piwi.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY014899; AAG42533.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   854 AA;  96722 MW;  509A1D39C0D1922C CRC64;

  Query Match             44.8%;  Score 2033;  DB 5;  Length 854;
  Best Local Similarity   47.4%;  Pred. No. 2.3e-147;
  Matches  409;  Conservative  143;  Mismatches  272;  Indels   38;  Gaps   12;

Qy         28 QQPGYIQPRPQ---PPPAEGE----LFGRGRQRGTAGGTAKSQGL--------QISAGFQ 72
              ::||   | |     || |      : |||| ||  | |    |          ::::
Db          3 RRPGEQAPPPMQQAAPPQEASSGMPVPGRGRSRG--GTTQPGPGAVPKQEPRGELTSQIA 60

Qy         73 ELSXAERG-----------GRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRL 121
              :|:    |              ||||  :|   ||   : ||::  |     : |  | |:|
Db         61 KLAIGASGEGRVGGAVQDRRRRRDEIELEPTTRPE-NFVKQAIAGDK---IALIANGFKL 116

Qy        122 TSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTEV 181
              ::| | |||| :|: | :   | | |||  |  |:||    |   ::   :| :||| :
Db        117 KTKPDWQLYQYRVDFEPEILNPRARFALLKGHSALLGKGLTLDMDTMYSLFKLAEKVTNL 176

Qy        182 FSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDI 241
              :: ::| :| | :|    | |  ||::||| ||::|::|:|||||:|   |||
Db        177 SAERKDGSNVNICVTHVATLNPMAPNTLHLYNVLFRRCLKMINMEQVGRNYYDPTAAIDI 236

Qy        242 PSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVSK 301
              | | :|||| |||||||  :|| :|:|||||||::|| : | : :::    :|:|:::|
Db        237 KQHGLQLWPGFVTSILQYEYDVMLLSDISHKVLRTQTVWEVMNDLFNKA-RGRFKEEITK 295

Qy        302 EXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQPV 361
              || :|||||| |||||||||:|  |  ||:   |  ||:::|::| | : | |: ||:
Db        296 LMIGQIVLTKYNNKTYRVDDIDFDTTPADTFETRSGP-VSYVDYFKKSYERVIHDVNQPM 354

Qy        362 LVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREV 421
              |||:||:|    |   ||| | ||||:||||:| || |||||||:  |||: |: | | :
Db        355 LVSRPKKREEKKGV--GPAYLPPELCFLTGLSDDMRADFNVMKDMGQHTRVGPQDRCRTL 412

Qy        422 GRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSKE 481
              |  :  |: |:  |  ||: ||   :  :||:|  ||: | || | |||  ||||::
Db        413 SGFIKKLSSNEEVKTYLDSWGMEFDKEQVKLTGRVLPPEKLFQRGKQFSYNPSNADWSRD 472

Qy        482 TRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEV-DDRTEA 540
              |||  |   | |:|| : ||||:       |::| :|   ||| :|   ::|: |||||
Db        473 TRGNALTDAKILNNWKIFYTRRDANRGQDFIKSLVRVANPMGMNVRGPEIVELPDDRTET 532

Qy        541 YLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAI 600
              | | || ::  |||||| :| :||||:||||||       | ||| :|:||| ||| :|::
Db        533 YTRSLQAQIAQDTQIVVVILPTNRKDRYDAIKKTCVVTHPCPSQVIVSRTLSKQQMLMSV 592

Qy        601 ATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFS 660
              |||||:|||||||:||||:|||  :||:||| |||     ||: |||||:|:  | :||
Db        593 ATKIAMQMNCKMGGDLWRVEIPLSNLMIIGIDSYHDSLTKGRSVLGFVASMNKSQTSFFS 652
```

```
Qy        661 RCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQF 720
Db        653 SCAFQHAQGEFGANLSTLMNNALKRYYQINEKFPERIIIFRDGVGDSQVNLVVDYELKQI 712

Qy        721 LDCL-KSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 779
Db        713 KDTLDKVYPQGTVHKLAVVVVKKRINNRFFANLRGGLSNPPPGTVIDDVVTKPHLYDYFI 772

Qy        780 VSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLA 839
Db        773 ISQSVRQGSVSPTSYNVVYDTTGLKPDHMQRLTYKLTHLYFNWPGTVRVPAPCMYAHKLA 832

Qy        840 FLVGQSIHREPNLSLSNRLYYL 861
Db        833 FLIGQSVHEVPNNRLADTLFYL 854


RESULT 8
Q9GPA7
ID   Q9GPA7      PRELIMINARY;      PRT;   808 AA.
AC   Q9GPA7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Seawi (Fragment).
OS   Strongylocentrotus purpuratus (Purple sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Strongylocentrotus.
OX   NCBI_TaxID=7668;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rodriguez A.J., Bonder E.M.;
RT   "Seawi - Cloning and Molecular Characterization of a Sea Urchin
RT   Homolog of Piwi.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY014900; AAG42534.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
FT   NON_TER     808    808
SQ   SEQUENCE   808 AA;  91366 MW;  19528F4B9D10474C CRC64;

  Query Match           40.5%;  Score 1835;  DB 5;  Length 808;
  Best Local Similarity   46.3%;  Pred. No. 3.5e-132;
  Matches  377;  Conservative  135;  Mismatches  264;  Indels   38;  Gaps   12;

Qy         28 QQPGYIQPRPQ---PPPAEGE----LFGRGRQRGTAGGTAKSQGL--------QISAGFQ 72
Db          3 RRPGEQAPPPMQQAAPPQEASSGMPVPGRGRSRG--GTTQPGPGAVPKQEPRGELTSQIA 60

Qy         73 ELSXAERG-----------GRRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRL 121
Db         61 KLAIGASGEGRVGGAVQDRRRRRDEIELEPTTRPE-NFVKQAIAGDK---IALIANGFKL 116

Qy        122 TSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTEV 181
Db        117 KTKPDWQLYQYRVDFEPEILNPRARFALLKGHSALLGKGLTLDMDTMYSLFKLAEKVTNL 176

Qy        182 FSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDI 241
Db        177 SAERKDGSNVNICVTHVATLNPMAPNTLHLYNVLFRRCLKMINMEQVGRNYYDPTAAIDI 236

Qy        242 PSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVSK 301
Db        237 KQHGLQLWPGFVTSILQYEYDVMLLSDISHKVLRTQTVWEVMNDLFNKA-RGRFKEEITK 295

Qy        302 EXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQPV 361
Db        296 LMIGQIVLTKYNNKTYRVDDIDFDTTPADTFETRSGP-VSYVDYFKKSYERVIHDVNQPM 354

Qy        362 LVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREV 421
Db        355 LVSRPKKREEKKGV--GPAYLPPELCFLTGLSDDMRADFNVMKDMGQHTRVGPQDRCRTL 412

Qy        422 GRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSKE 481
Db        413 SGFIKKLSSNEEVKTYLDSWGMEFDKEQVKLTGRVLPPEKLFQRGKQFSYNPSNADWSRD 472

Qy        482 TRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKAIMIEV-DDRTEA 540
Db        473 TRGNALTDAKILNNWKIFYTRRDANRGQDFIKSLVRVANPMGMNVRGPEIVELPDDRTET 532

Qy        541 YLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAI 600
Db        533 YTRSLQAQIAQDTQIVVVILPTNRKDRYDAIKKTCVVTHPCPSQVIVSRTLSKQQMLMSV 592

Qy        601 ATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFS 660
Db        593 ATKIAMQMNCKMGGDLWRVEIPLSNLMIIGIDSYHDSLTKGRSVLGFVASMNKSQTSFFS 652

Qy        661 RCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQF 720
Db        653 SCAFQHAQGEFGANLSTLMNNALKRYYQINEKFPERIIIFRDGVGDSQVNLVVDYELKQI 712

Qy        721 LDCL-KSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFI 779
Db        713 KDTLDKVYPQGTVHKLAVVVVKKRINNRFFANLRGGLSNPPPGTVIDDVVTKPHLYDYFI 772

Qy        780 VSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTY 813
Db        773 ISQSVRQGSVSPTSYNVVYDTTGLKPDHMQRLTY 806


RESULT 9
Q9JMB6
ID   Q9JMB6      PRELIMINARY;      PRT;   971 AA.
AC   Q9JMB6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   MILI (Miwi like).
GN   PIWIL2 OR PIWIL1L OR MILI.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Miyagawa S.K., Kimura T., Nakano T.;
RT   "Molecular Cloning and characterization of Piwi family genes.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB032605; BAA93706.1; -.
DR   MGD; MGI:1930036; Piwil2.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   971 AA;  109488 MW;  01E143C6513310FB CRC64;

  Query Match           38.4%;  Score 1742;  DB 11;  Length 971;
  Best Local Similarity   42.5%;  Pred. No. 6.5e-125;
  Matches  324;  Conservative  161;  Mismatches  268;  Indels   10;  Gaps    9;

Qy        104 KTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAF 163
Db        214 KQGSKGTPQSLGLNLIKIQCHNE-AVYQYHVTFSPSVECKSMRFGMLKDHQSVTGNVTAF 272

Qy        164 DGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIM 223
Db        273 DGSILYLPVKLQQVVELKSQRKTDDAEISIKIQLTKILEPCSDLCIPFYNVVFRRVMKLL 332

Qy        224 NLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFM 283
Db        333 DMKLVGRNFYDPTSAMVLQQHRLQIWPGYAASIRRTDGGLFLLADVSHKVIRNDSVLDVM 392
```

```
Qy        284 FNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFL 343
Db        393 HAIYQQNKEH-FQDECSKLLVGSIVITRYNNRTYRIDDVDWNKTPKDSFVMSDGKEITFL 451

Qy        344 EYYRKQYNQEITDLKQPVLVSQP-KRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNV 402
Db        452 EYYSKNYGITVKEDDQPLLIHRPSERQNNHGMLLKGEILLLPELSFMTGIPEKMKKDFRA 511

Qy        403 MKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKI 462
Db        512 MKDLTQQINLSPKQHHGALECLLQRISQNETASNELTRWGLSLHKDVHKIEGRLLPMERI 571

Qy        463 HQGGKTFDYNPQFADWSKE-TRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPA 521
Db        572 NLRNTSF-VTSEDLNWVKEVTRDASILTI-PMHFWALFYPKRAMDQARELVNMLEKIAGP 629

Qy        522 MGMQMRKAIMIEV-DDRTEAYLRVLQQ--KVTADTQIVVCLLSSNRKDKYDAIKKYLCTD 578
Db        630 IGMRISPPAWVELKDDRIETYIRTIQSLLGVEGKIQMVVCIIMGTRDDLYGAIKKLCCVQ 689

Qy        579 CPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMT 638
Db        690 SPVPSQVINVRTIGQPTRLRSVAQKILLQMNCKLGGELWGVDIPLKQLMVIGMDVYHDPS 749

Qy        639 AGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRII 698
Db        750 RGMRSVVGFVASINLTLTKWYSRVVFQMPHQEIVDSLKLCLVGSLKKYYEVNHCLPEKIV 809

Qy        699 VYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQN 758
Db        810 VYRDGVSDGQLKTVANYEIPQLQKCFEAFD-NYHPKMVVFVVQKKISTNLYLAAPDHFVT 868

Qy        759 PLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHI 818
Db        869 PSPGTVVDHTITSCEWVDFYLLAHHVRQGCGIPTHYICVLNTANLSPDHMQRLTFKLCHM 928

Qy        819 YYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 861
Db        929 YWNWPGTIRVPAPCKYAHKLAFLSGQILHHEPAIQLCGNLFFL 971


RESULT 10
Q8TC59
ID   Q8TC59      PRELIMINARY;      PRT;   973 AA.
AC   Q8TC59;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Similar to piwi like homolog 1 (Drosophila)-like.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Strausberg R.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC025995; AAH25995.1; -.
SQ   SEQUENCE   973 AA;  109848 MW;  C44398136B144CA0 CRC64;

  Query Match           38.2%;  Score 1730;  DB 4;  Length 973;
  Best Local Similarity   42.3%;  Pred. No. 5.5e-124;
  Matches  322;  Conservative  156;  Mismatches  276;  Indels    8;  Gaps    7;

Qy        104 KTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAF 163
Db        216 KQGSKGTPQSLGLNLVKIQCHNE-AVYQYHVTFSPNVECKSMRFGMLKDHQAVTGNVTAF 274

Qy        164 DGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIM 223
Db        275 DGSILYLPVKLQQVLELKSQRKTDSAEISIKIQMTKILEPCSDLCIPFYNVVFRRVMKLL 334
```

```
Qy        224 NLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFM 283
Db        335 DMKLVGRNFYDPTSAMVLQQHRLQIWPGYAASIRRTDGGLFLLADVSHKVIRNDCVLDVM 394

Qy        284 FNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFL 343
Db        395 HAIYQQNKEH-FQDECTKLLVGNIVITRYNNRTYRIDDVDWNKTPKDSFTMSDGKEITFL 453

Qy        344 EYYRKQYNQEITDLKQPVLVSQPKRRR-GPGGTLPGPAMLIPELCYLTGLTDKMRNDFNV 402
Db        454 EYYSKNYGITVKEEDQPLLIHRPSERQDNHGMLLKGEILLLPELSFMTGIPEKMKKDFRA 513

Qy        403 MKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKI 462
Db        514 MKDLAQQINLSPKQHHSALECLLQRIAKNEAATNELMRWGLRLQKDVHKIEGRVLPMERI 573

Qy        463 HQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAM 522
Db        574 NLKNTSF-ITSQELNWVKEVTRDPSILTIPMHFWALFYPKRAMDQARELVNMLEKIAGPI 632

Qy        523 GMQMRKAIMIEV-DDRTEAYLRVLQQKVTAD--TQIVVCLLSSNRKDKYDAIKKYLCTDC 579
Db        633 GMRMSPPAWVELKDDRIETYVRTIQSTLGAEGKIQMVVCIIMGPRDDLYGAIKKLCCVQS 692

Qy        580 PTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTA 639
Db        693 PVPSQVVNVRTIGQPTRLRSVAQKILLQINCKLGGELWGVDIPLKQLMVIGMDVYHDPSR 752

Qy        640 GRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIV 699
Db        753 GMRSVVGFVASINLTLTKWYSRVVFQMPHQEIVDSLKLCLVGSLKKFYEVNHCLPEKIVV 812

Qy        700 YRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNP 759
Db        813 YRDGVSDGQLKTVANYEIPQLQKCFEAF-ENYQPKMVVFVVQKKISTNLYLAAPQNFVTP 871

Qy        760 LPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIY 819
Db        872 TPGTVVDHTITSCEWVDFYLLAHHVRQGCGIPTHYVCVLNTANLSPDHMQRLTFKLCHMY 931

Qy        820 YNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 861
Db        932 WNWPGTIRVPAPCKYAHKLAFLSGHILHHEPAIQLCENLFFL 973


RESULT 11
O76922
ID   O76922      PRELIMINARY;      PRT;   866 AA.
AC   O76922;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   AUB protein.
GN   AUB OR STING OR CG6137.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
```

```
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON S;
RX   MEDLINE=96262192; PubMed=8676882;
RA   Schmidt A., Hollmann M., Schaefer U.;
RT   "A newly identified Minute locus, M(2)32D, encodes the ribosomal
RT   protein L9 in Drosophila melanogaster.";
RL   Mol. Gen. Genet. 251:381-387(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON S;
RA   Schmidt A., Palumbo G., Bozzetti M.P., Tritto P., Pimpinelli S.,
RA   Schaefer U.;
RT   "Genetic and molecular characterization of sting, a gene involved in
RT   crystal formation and meiotic drive in Drosophila melanogaster.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Tsang G., Brokstein P., Frise E., Harvey D., Evans-Holm M.,
RA   Lewis S.E., Suh C., Rubin G.M.;
RT   "Full Length Drosophila melanogaster cDNA sequence.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003630; AAF53046.1; -.
DR   EMBL; X94613; CAA64320.1; -.
DR   EMBL; AF145680; AAD38655.1; -.
DR   FlyBase; FBgn0000146; aub.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   866 AA;  98558 MW;  8F84CF1B0B89103B CRC64;

  Query Match             32.3%;  Score 1464;  DB 5;  Length 866;
  Best Local Similarity   36.6%;  Pred. No. 1.3e-103;
  Matches  323;  Conservative  165;  Mismatches  340;  Indels   54;  Gaps   17;

Qy          9 ARGRARGQETAQLVGSTASQQPGYIQPRPQPPPAEGELFGRGRQRG--TAGGTAKSQGLQ 66
              |||| ||            ::|  ::      |: |:        |   :||       |
Db         10 ARGRGRG-----------RKPNNVEANRGFAPSLGQKSDPSHSEGNQASGGNGGGGDAQ 57

Qy         67 ISAGFQELSXA----------ERGG--RRRDFHDLGVNTRQNLDHVKESKTGSSGIIVRL 114
              :    :: | :           ||    ||    ||  :    :     || |  |  : :
Db         58 VGPSIEKSSLSAVQMHKSEGDPRGSVRGRRLITDLVYSRPPGM----TSKKGVVGTHITV 113

Qy        115 STNHFRLTSRPQWALYQYHIDYNPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLP--- 171
              |:|::  || | :||| :|: | :|| ||| : |::|: ::|  : |||| :|
Db        114 QANYFKVLKRPNWTIYQYRVDFTPDVEATRLRRSFLYEHKGILGG-YIFDGTNMFCINQF 172

Qy        172 KRLQQK--VTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIG 229
              | :|    | |: :|:| ||:: | |     :  |     |  |:| || :: ::|: :
Db        173 KAVQDSPYVLELVTKSRAGENIEIKIKAVGSVQSTDAEQFQVLNLILRRAMEGLDLKLVS 232

Qy        230 RNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQ 289
              | ||:|   |:: : |: :|||: ||| |:|| |:||::: |||:|:||: : : :
Db        233 RYYYDPQAKINLENFRMQLWPGYQTSIRQHENDILLCSEICHKVMRTETLYNILSDAIRD 292

Qy        290 TEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQ 349
              :::  :|    :  :|:|:|| || ||||:||:|:   |   ||  ||  :|:::||:|:
Db        293 SDD--YQSTFKRAVMGMVILTDYNNKTYRIDDVDFQSTPLCKFKTNDGE-ISYVDYYKKR 349

Qy        350 YNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVH 409
              ||  | |||||:::|:|  :  ||      |:|||   ||:|| || ||  :: :: |
Db        350 YNIIIRDLKQPLVMSRPTDKNIRGGN-DQAIMIIPELARATGMTDAMRADFRTLRAMSEH 408

Qy        410 TRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEKIHQGGKTF 469
              ||| |::|   :      :          |: | :  || |:    |:|  |||  | :
Db        409 TRLNPDRRIERLRMFNKRLKSCKQSVETLKSWNIELDSALVEIPARVLPPEKILFGNQKI 468

Qy        470 DYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRKA 529
                  |||: | |    :    :: | :|   ||       :|   :   :| | :
Db        469 FVCDARADWTNEFRTCSMFKNVHINRWYVITPSRNLRETQEFVQMCIRTASSMKMNICNP 528

Qy        530 IMIEV-DDRTEAYLRVLQQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVA 588
              |  |: |||   | : :       | |||: :: |  ::||  |||  | | | ||| |
Db        529 IYEEIPDDRNGTYSQAIDNAAANDPQIVMVVMRSPNEEKYSCIKKRTCVDRPVPSQVVTL 588

Qy        589 RTLGKQQT----VMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSI 644
              : :  :|      :|:|||||: :||| |: |  |:| |||  :| || |  |       :|
Db        589 KVIAPRQQKPTGLMSIATKVVIQMNAKLMGAPWQVVIPLHGLMTVGFDVCHSPKNKNKSY 648

Qy        645 AGFVASINEGMT-RWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDG 703
              |||::::  : |:||      :|||| : | : |  |||::   :  :| ||: :|||
Db        649 GAFVATMDQKESFRYFSTVNEHIKGQELSEQMSVNMACALRSYQEQHRSLPERILFFRDG 708

Qy        704 VGDGQLKTLVNYEVPQFLDCL----KSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQNP 759
              ||||||  :|| ||    | |    || |:    | | |:| ||:|:|:|  :| |  ||
Db        709 VGDGQLYQVVNSEVNTLKDRLDEIYKSAGKQEGCRMTFIIVSKRINSRYF--TGHR--NP 764

Qy        760 LPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIY 819
              :||||:|  :| || ||||:|||||| |:|||| |||| || ||  | :| |:||: |:|
Db        765 VPGTVVDDVITLPERYDFFLVSQAVRIGTVSPTSYNVISDNMGLNADKLQMLSYKMTHMY 824

Qy        820 YNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 861
              ||: | ||||| | |||||||||| :|||:| |:  | |:||:|
Db        825 YNYSGTIRVPAVCHYAHKLAFLVAESINRAPSAGLQNQLYFL 866


RESULT 12
Q9VKM1
ID   Q9VKM1      PRELIMINARY;      PRT;   843 AA.
AC   Q9VKM1; O96674; O96675;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Piwi protein.
GN   PIWI OR CG6122.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   STRAIN=OREGON-R, AND CANTON-S;
RX   MEDLINE=99069219; PubMed=9851978;
RA   Cox D.N., Chao A., Baker J., Chang L., Qiao D., Lin H.;
RT   "A novel class of evolutionarily conserved genes defined by piwi are
```

```
RT   essential for stem cell self-renewal.";
RL   Genes Dev. 12:3715-3727(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: MEDIATES A SOMATIC SIGNALING MECHANISM REQUIRED FOR THE
CC       ASYMMETRIC DIVISION OF GERM-LINE STEM CELLS TO PRODUCE AND
CC       MAINTAIN A DAUGHTER GERM-LINE STEM CELL. IT IS NOT ESSENTIAL FOR
CC       THE FURTHER DIFFERENTIATION OF THE COMMITTED DAUGHTER CELL.
CC   -!- TISSUE SPECIFICITY: EXPRESSED SOMATICALLY IN THE OVARIOLE TERMINAL
CC       FILAMENT AND EPITHELIAL SHEATH CELLS. EXPRESSED IN THE GERMARIUM
CC       AT LOW LEVELS DURING OOGENESIS STAGES 1-6, AT A LOWER LEVEL DURING
CC       STAGES 7-9, STRONGLY AT STAGE 10 AND EVENTUALLY ACCUMULATING IN
CC       EARLY EMBRYOS.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED BOTH MATERNALLY AND ZYGOTICALLY.
CC   -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
DR   EMBL; AF104354; AAD08704.1; -.
DR   EMBL; AF104355; AAD08705.1; -.
DR   EMBL; AE003630; AAF53043.1; -.
DR   FlyBase; FBgn0004872; piwi.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
KW   Developmental protein.
FT   CONFLICT     67     67       T -> P (IN REF. 1; AAD08704).
FT   CONFLICT     84     84       A -> V (IN REF. 1; AAD08704).
FT   CONFLICT    122    124       EPS -> VPT (IN REF. 1; AAD08704).
FT   CONFLICT    576    577       NR -> KPY (IN REF. 1; AAD08705).
SQ   SEQUENCE   843 AA;  97177 MW;  9B4A95E688E9D9B7 CRC64;

  Query Match             31.3%;  Score 1417;  DB 5;  Length 843;
  Best Local Similarity   35.9%;  Pred. No. 5e-100;
  Matches  309;  Conservative  170;  Mismatches  313;  Indels   68;  Gaps   19;

Qy         48 GRGRQR-------GTAGGTAKSQGLQISAG--------------------FQELSXAERG 80
              ||||:|        |: |:     :::  |                       :|    | |
Db          6 GRGRRRPLNEDDSSTSRGSGDGPRVKVFRGSSSGDPRADPRIEASRERRALEEAPRREGG 65

Qy         81 GRRR----DFHDLGVNTRQNLDHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDY 136
                 |     | :|  :|||         || |: |: | | || ||| ::|:| :  ||::: 
Db         66 PTERKPWGDQYDY-LNTR---PAELVSKKGTDGVPVMLQTNFFRLKTKPEWRIVHYHVEF 121

Qy        137 NPLMEARRLRSALLFQHEDLIGKCHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITIT 196
               | :|  |:|  :|  | :|:|  : |||  ||  :: :|:::| :  |::   : :|:| 
Db        122 EPSIENPRVRMGVLSNHANLLGSGYLFDGLQLFTTRKFEQEITVLSGKSKLDIEYKISIK 181

Qy        197 LTNELPPTSPTCLQFYNIIFRRLLKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSI 256
                 :    |  ||  |:| || :| :||: :||| ::|   |:|   :: :|||: |||
Db        182 FVGFISCAEPRFLQVLNLILRRSMKGLNLELVGRNLFDPRAKIEIREFKMELWPGYETSI 241

Qy        257 LQYENSIMLCTDVSHKVLRSETVLDFMFNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKT 316
               |:|  |:| |:::|||:|:||: | |    |    |  |::|    : |:||| || :|
Db        242 RQHEKDILLGTEITHKVMRTETIYDIMRRCSHNPARH--QDEVRVNVLDLIVLTDYNNRT 299

Qy        317 YRVDDIDWDQNPKSTFKKADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTL 376
               ||::|:|: | |||||     |  :||:|||  :||  | |  ||:|:|:  | :
Db        300 YRINDVDFGQTPKSTF-SCKGRDISFVEYYLTKYNIRIRDHNQPLLISK-NRDKALKTNA 357

Qy        377 PGPAMLIPELCYLTGLTDKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQ- 435
                 :|||||| :|||  :||::| :|: :: :||: |:||     ||  : |:  |
Db        358 SELVVLIPELCRVTGLNAEMRSNFQLMRAMSSYTRMNPKQR---TDRLRAFNHRLQNTPE 414

Qy        436 --RELRDWGLSFDSNLLSFSGRILQTEKI--HQGGKTFDYNPQFADWSKETRGAPLISVK 491
                 : |||| :  | |:    |||:  : |  | |       |   ||| :  |   :::
Db        415 SVKVLRDWNMELDKNVTEVQGRIIGQQNIVFHNGKVPAGEN---ADWQRHFRDQRMLTTP 471

Qy        492 P--LDNWLLIYTRRNYEAANSLIQNLFKVTPAMGMQMRK-AIMIEVDDRTEAYLRVLQQK 548
               || | :|  :||     :|: :|::    ||:::|     |  |||| |:| :
Db        472 SDGLDRWAVIAPQRNSHELRTLLDSLYRAASGMGLRIRSPQEFIIYDDRTGTYVRAMDDC 531

Qy        549 VTADTQIVVCLLSSNRKDKYDAIKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQM 608
               | :| ::::||: ::  ::| :|||    |   |:| |  :|  | :::|:|||||:|:
Db        532 VRSDPKLILCLVPNDNAERYSSIKKRGYVDRAVPTQVVTLKTT-KNRSLMSIATKIAIQL 590

Qy        609 NCKMGGELWRVDIPLKLVMIVGIDCYHDMTAGRRSIAGFVASIN-EGMTRWFSRCIFQDR 667
               |||:|   | :::||  :| :| |        :|:    :||:: :  : :||
Db        591 NCKLGYTPWMIELPLSGLMTIGFDIAKSTRDRKRAYGALIASMDLQQNSTYFSTVTECSA 650

Qy        668 GQELVDGLKVCLQAALRAWNSCNEYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSI 727
               | : |   :  ||| :  :|||||       ||| |||::|  :| || |||::||
Db        651 FDVLANTLWPMIAKALRQYQHEHRKLPSRIVFYRDGVSSGSLKQLFEFEVKDIIEKLKT- 709

Qy        728 GRGY------NPRXTVIVVKKRVNTRFFAQSGGRLQNPLPGTVIDVEVTRPEWYDFFIVS 781
               |       |:   ||| : :|||||       ||| |||::|  :| || |||::||
Db        710 --EYARVQLSPPQLAYIVVTRSMNTRFFLNG----QNPPPGTIVDDVITLPERYDFYLVS 763

Qy        782 QAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFL 841
               | || |:|||| |||:| : || |: :|:||||:||:|||| |  |||| |||| ||| |
Db        764 QQVRQGTVSPTSYNVLYSSMGLSPEKMQKLTYKMCHLYYNWSGTTRVPAVCQYAKKLATL 823

Qy        842 VGQSIHREPNLSLSNRLYYL 861
               || ::|  |  :|  : |||
Db        824 VGTNLHSIPQNALEKKFYYL 843


RESULT 13
Q99MV6
ID   Q99MV6      PRELIMINARY;      PRT;   580 AA.
AC   Q99MV6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   PIWIL2.
GN   PIWIL1L OR PIWIL2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE=21175748; PubMed=11279525;
RA   Wang P.J., McCarrey J.R., Yang F., Page D.C.;
RT   "An abundance of X-linked genes expressed in spermatogonia.";
RL   Nat. Genet. 27:422-426(2001).
DR   EMBL; AF285586; AAK31965.1; -.
DR   MGD; MGI:1930036; Piwil2.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   580 AA;  66549 MW;  A4EA9DE8626FA136 CRC64;

  Query Match             30.1%;  Score 1365.5;  DB 11;  Length 580;
  Best Local Similarity   43.7%;  Pred. No. 2.6e-96;
  Matches  255;  Conservative  117;  Mismatches  203;  Indels    9;  Gaps    8;

Qy        283 MFNFYHQTEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFKKADGSGVSF 342
Db          1 MHAIYQQNKEH-FQDECSKLLVGSIVITRYNNRTYRIDDVDWNKTPKDSFVMSDGKEITF 59

Qy        343 LEYYRKQYNQEITDLKQPVLVSQP-KRRRGPGGTLPGPAMLIPELCYLTGLTDKMRNDFN 401
Db         60 LEYYSKNYGITVKEDDQPLLIHRPSERQNNHGMLLKGEILLLPELSFMTGIPEKMKKDFR 119

Qy        402 VMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFSGRILQTEK 461
Db        120 AMKDLTQQINLSPKQHHGALECLLQRISQNETASNELTRWGLSLHKDVHKIEGRLLPMER 179

Qy        462 IHQGGKTFDYNPQFADWSKE-TRGAPLISVKPLDNWLLIYTRRNYEAANSLIQNLFKVTP 520
Db        180 INLRNTSF-VTSEGLNWVKEVTRDASILTI-PMHFWALFYPKRAMDQARELVNMLEKIAG 237

Qy        521 AMGMQMRKAIMIEV-DDRTEAYLRVLQQ--KVTADTQIVVCLLSSNRKDKYDAIKKYLCT 577
Db        238 PIGMRTSPPAWVELKDDRIETYIRTIQSLLGVEGKIQMVVCIIMGTRDDLYGAIKKLCCV 297

Qy        578 DCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVGIDCYHDM 637
Db        298 QSPVPSQVINVRTIGQPTRLRSVAQKILLQMNCKLGGELWGVDIPLKQLMVIGMDVYHDP 357

Qy        638 TAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCNEYMPSRI 697
Db        358 SRGMRSVVGFVASINLTLTKWYSRVVFQMPHQEIVDSLKLCLVGSLKKYYEVNHCLPEKI 417

Qy        698 IVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFAQSGGRLQ 757
Db        418 VVYRDGVSDGQLKTVANYEIPQLQKCFEAFD-NYHPKMVVFVVQKKISTNLYLAAPDHFV 476

Qy        758 NPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQRLTYKLCH 817
Db        477 TPSPGTVVDHTITSCEWVDFYLLAHHVRQGCGIPTHYICVLNTANLSPDHMQRLTFKLCH 536

Qy        818 IYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 861
Db        537 MYWNWPGTIRVPAPCKYAHKLAFLSGQILHHEPAIQLCGNLFFL 580


RESULT 14
P90786
ID   P90786      PRELIMINARY;      PRT;   824 AA.
AC   P90786;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   D2030.6 protein.
GN   D2030.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wilkinson J.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z73906; CAA98113.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   824 AA;  93844 MW;  970263CF6FEDBCEE CRC64;

  Query Match             28.5%;  Score 1291.5;  DB 5;  Length 824;
  Best Local Similarity   33.8%;  Pred. No. 2.2e-90;
  Matches  289;  Conservative  174;  Mismatches  317;  Indels   75;  Gaps   19;

Qy         48 GRGRQRGTAGGTAKSQGLQISAGFQELSXAERGGRRRDF-----HDLGVNTRQNLDHVKE 102
Db          4 GSGRGRGRGSGSNNS----------------GGKDQKYLGTIQPDLFI--RQQ----GQ 40

Qy        103 SKTGSSGIIVRLSTNHFRL-TSRPQWALYQYHIDYNPLMEARRLRSALLFQHE--DLIGK 159
Db         41 SKTGSSGQPQKCFANFIPIEMTQSDYSIYQYHVEFEPTVDSKANREKMLRDNNVTDEIGH 100

Qy        160 CHAFDGTILFLPKRLQQKVTEVFSKTRNGEDVRITITLTNELPPTSPTCLQFYNIIFRRL 219
Db        101 HFVFDGMILYLKEEWEQNQMIEVQHPIDRSLICIRFKQTNRFLVDDPQTINIFNTIIRRS 160

Qy        220 LKIMNLQQIGRNYYNPNDPIDIPSHRLVIWPGFTTSILQYENSIMLCTDVSHKVLRSETV 279
Db        161 FDALQLTQLGRNYFNWGDSRAVPDYNMSILPGYETAIRMYEENFMLCVENRFKMVREES- 219

Qy        280 LDFMFNFYH------QTEEHKFQEQVSKEXIGLVVLTKYNIKTYRVDDIDWDQNPKSTFK 333
Db        220 ---MYILFHKELRSCQNNPQRVQEKMNEMYGGTTIITRYNNKLHRYTRLDYSISPLSEFV 276

Qy        334 KADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRRGPGGTLPGPAMLIPELCYLTGLT 393
Db        277 K-DGQSITLKEYFKNQYGIEITVDDQPIIISEGKPKQ--PGEPPQVSYIVPELCFPTGLT 333

Qy        394 DKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFS 453
Db        334 DEMRKDFKMMKEIAKHTRMSPQQRLVESRKLIVDLSKNEKVMECFKYWGISLGQDLANVQ 393

Qy        454 GRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEA--ANSL 511
Db        394 ARVLKSEPL-QGKKT--YEGKQAEWARGVKECGIYRGSNMTNWIVIGPGSGNSGLLSQKF 450

Qy        512 IQNLFKVTPAMGMQMRKAIMIEV-----DDRTEAYLRVLQQKVTADTQIVVCLLSSNRKD 566
Db        451 IEEARRLGKILQVQLGEPMCVPIRGISPNDYLEGVKGAIKQVDGEDIHMLVVMLADDNKT 510

Qy        567 KYDAIKKYLCTDCPTPSQCVVARTL-------GKQQTVMAIATKIALQMNCKMGGELWRV 619
Db        511 RYDSLKKFLCVECPIPNQCVNLRTLAGKSKDGGENKNLGSIVLKIVLQMICKTGGALWKV 570

Qy        620 DIPLKLVMIVGIDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCL 679
Db        571 NIPLKNTMIVGYDLYHDSTLKGKTVGACVSTTSNDFTQFYSQTRPHENPTQLGNNLTHFV 630

Qy        680 QAALRAWNSCNEY-MPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGR--------- 729
Db        631 RKALKQYYDSNDQTLPSRLILYRDGAGDGQIPYIKNTEVKLVRDACDAVTDKAAELSNKV 690

Qy        730 GYNPRXTVIVVKKRVNTRFFAQSGGRLQ---NPLPGTVIDVEVTRPEWYDFFIVSQAVRS 786
```

```
                :   |:| |||| |   | |  |    || ||||:|  |||||  ||::| | |
Db        691 QEKIKLAFIIVTKRVNMRILKQ-GSSLDNAINPQPGTVVDTTVTRPERMDFYLVPQFVNQ 749

Qy        787 GSVSPTHYNVIYDNSGLKPDHIQRLTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSI 846
              |:|:|  ||:|:|:: |'||  |:| :||||:|||| | :|||||||||||||||  ||:
Db        750 GTVTPVSYNIIHDDTDLGPDKHQQLAFKLCHLYYNWQGTVRVPAPCQYAHKLAFLTAQSL 809

Qy        847 HREPNLSLSNRLYYL 861
              | : |  | ::|::|
Db        810 HDDANGCLRDKLFFL 824


RESULT 15
Q96SW6
ID   Q96SW6      PRELIMINARY;      PRT;   530 AA.
AC   Q96SW6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   CDNA FLJ14591 fis, clone NT2RM4002034, weakly similar to Homo
DE   sapiens hiwi mRNA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA   Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA   Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA   Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA   Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA   Ninomiya K., Iwayanagi T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK027497; BAB55155.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02171; Piwi; 1.
SQ   SEQUENCE   530 AA;  60620 MW;  BA4A8AADD809B32F CRC64;

  Query Match             26.9%;  Score 1220;  DB 4;  Length 530;
  Best Local Similarity   43.5%;  Pred. No. 3.4e-85;
  Matches  231;  Conservative   98;  Mismatches  196;  Indels    6;  Gaps    5;

Qy        335 ADGSGVSFLEYYRKQYNQEITDLKQPVLVSQPKRRR-GPGGTLPGPAMLIPELCYLTGLT 393
              :||  ::|||||| | |   : :  ||:|: :|  |:   |  | |  :|:||| ::||:
Db          2 SDGKEITFLEYYSKNYGITVKEEDQPLLIHRPSERQDNHGMLLKGEILLLPELSFMTGIP 61

Qy        394 DKMRNDFNVMKDLAVHTRLTPEQRQREVGRLIDYIHKNDNVQRELRDWGLSFDSNLLSFS 453
              :||: ||  |||||    |:|:|    :  |:  | ||:    ||  |||    ::
Db         62 EKMKKDFRAMKDLAQQINLSPKQHHSALECLLQRIAKNEAATNELMRWGLRLQKDVHKIE 121

Qy        454 GRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAANSLIQ 513
              ||:|  |:|:    :|    |  :| ||    | |   |:  | | | :|  : |  |:
Db        122 GRVLPMERINLKNTSF-ITSQELNWVKEVTRDPSILTIPMHFWALFYPKRAMDQARELVN 180

Qy        514 NLFKVTPAMGMQMRKAIMIEV-DDRTEAYLRVLQQKVTAD--TQIVVCLLSSNRKDKYDA 570
               | |:   :||:|     :|: ||| | |:| :|  : |:   |:|||::   | | | |
Db        181 MLEKIAGPIGMRMSPPAWVELKDDRIETYVRTIQSTLGAEGKIQMVVCIIMGPRDDLYGA 240

Qy        571 IKKYLCTDCPTPSQCVVARTLGKQQTVMAIATKIALQMNCKMGGELWRVDIPLKLVMIVG 630
              |||  |   | ||| |  ||:|:   : ::| || ||:|||:|||| |||||| :|::|
Db        241 IKKLCCVQSPVPSQVVNVRTIGQPTRLRSVAQKILLQINCKLGGELWGVDIPLKQLMVIG 300

Qy        631 IDCYHDMTAGRRSIAGFVASINEGMTRWFSRCIFQDRGQELVDGLKVCLQAALRAWNSCN 690
              :| ||| : | ||: |||||||  :|:|:|| :||   ||:|| ||:||  :|: :   |
Db        301 MDVYHDPSRGMRSVVGFVASINLTLTKWYSRVVFQMPHQEIVDSLKLCLVGSLKKFYEVN 360

Qy        691 EYMPSRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRXTVIVVKKRVNTRFFA 750
              :| :|:|||||| ||||||: |||:||   | ::    | |:  | ||:|:::|  :
Db        361 HCLPEKIVVYRDGVSDGQLKTVANYEIPQLQKCFEAF-ENYQPKMVVFVVQKKISTNLYL 419

Qy        751 QSGGRLQNPLPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDNSGLKPDHIQR 810
              :      | ||||:|  :|  || ||:::  || |   ||||  : : : | |||:||
Db        420 AAPQNFVTPTPGTVVDHTITSCEWVDFYLLAHHVRQGCGIPTHYVCVLNTANLSPDHMQR 479

Qy        811 LTYKLCHIYYNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSLSNRLYYL 861
              ||:||||:|:|||| |||||||:|||||||| |  :|  | : |   |::|
Db        480 LTFKLCHMYWNWPGTIRVPAPCKYAHKLAFLSGHILHHGPAIQLCENLFFL 530


Search completed: July 22, 2003, 07:04:50
Job time : 35.5542 secs
```